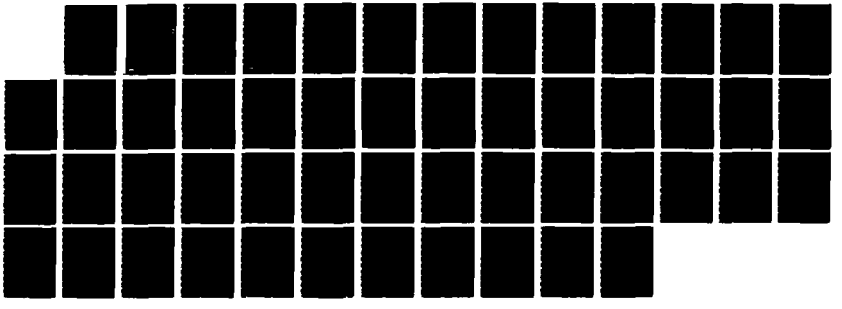
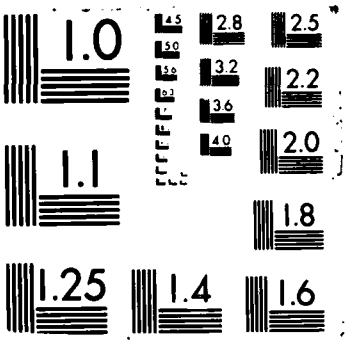
AD-A181 422 KNOWLEDGE REPRESENTATION AND NATURAL-LANGUAGE SEMANTICS 1/1
(U) SRI INTERNATIONAL MENLO PARK CA R C MOORE
07 NOV 86 AFOSR-TR-87-0784 F49620-85-K-0012

UNCLASSIFIED F/G 12/5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3

(2)

KNOWLEDGE REPRESENTATION AND NATURAL-LANGUAGE SEMANTICS

Final Report

Covering the Period June 1, 1985 to July 30, 1986

November 1986

By: Robert C. Moore, Staff Scientist

Contributing Author:

David J. Israel, Senior Computer Scientist

Artificial Intelligence Center
Computer and Information Sciences Division

Prepared for:

Air Force Office of Scientific Research
Building 410
Bolling Air Force Base
Washington, D.C. 20332

Attention: Dr. Robert Buchal

Contract No. F49620-85-K-0012

SRI Project 8666

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025-3493
(415) 326-6200
TWX: 910-373-2046
Telex: 334486

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

A181422

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
N/A

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release, distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
SRI Project ECU-8666

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 87-0784

6a. NAME OF PERFORMING ORGANIZATION
SRI International

6b. OFFICE SYMBOL (If applicable)
NM

7a. NAME OF MONITORING ORGANIZATION
U.S. Air Force Office of Scientific Research
NM

6c. ADDRESS (City, State, and ZIP Code)
333 Ravenswood Avenue
Menlo Park, CA 94025

7b. ADDRESS (City, State, and ZIP Code)
Bolling Air Force Base
Washington, D.C. 20332

8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Air Force Office of Scientific Research

8b. OFFICE SYMBOL (If applicable)
nm

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620-85-K-0012

8c. ADDRESS (City, State, and ZIP Code)
Bolling Air Force Base
Washington, D.C. 20332

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61103F	2304	A5	

11. TITLE (Include Security Classification)
Knowledge Representation and Natural-Language Semantics

12. PERSONAL AUTHOR(S)
Robert C. Moore

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 6-1-85 TO 7-30-86	1986 November 7	46

16. SUPPLEMENTARY NOTATION
Contributing Author: David J. Israel

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Artificial Intelligence; Knowledge Representation; Theories of Knowledge and Action; Natural-Language Semantics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report summarizes a year of research on a project both to develop a richer and more systematic theory of the role of beliefs in the production of rational action and to develop formalisms, suitable for manipulation by computer, adequate for representing and reasoning about the concepts involved in the theoretical analysis. Specific topics on which progress was made include: an account of the dependence of successful action on belief and knowledge, and a semantical analysis of adverbial modifiers and event sentences

Keywords:

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL
~~Dr. Robert Buchal~~ Vince Sigillito

22b. TELEPHONE (Include Area Code)
(202) 767-4939

22c. OFFICE SYMBOL
nm

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

SRI International

KNOWLEDGE REPRESENTATION AND NATURAL-LANGUAGE SEMANTICS

Final Report

Covering the Period June 1, 1985 to July 30, 1986

November 1986

By: Robert C. Moore, Staff Scientist

Contributing Author:

David J. Israel, Senior Computer Scientist

Artificial Intelligence Center
Computer and Information Sciences Division

Prepared for:

Air Force Office of Scientific Research
Building 410
Bolling Air Force Base
Washington, D.C. 20332

Attention: Dr. Robert Buchal

Contract No. F49620-85-K-0012

SRI Project 8666

Approved:

Stanley J. Rosenschein, Director
Artificial Intelligence Center

Donald L. Nielson, Vice President and Director
Computer and Information Sciences Division

SRI International

Contents

1 OBJECTIVES OF THE RESEARCH REPORT

The development of models and theories of rational action is crucial to the design and construction of the next generation of intelligent machines. The form of all such models can be characterized as follows: to make rigorous and systematic the common sense picture of rational actions as those that would tend to promote the agents' desires and goals if their beliefs were true. A schematic version of the dependence of rational action on beliefs and desires is often given in the form of the *practical syllogism*:

- All things considered, I want it to be the case that Φ.
- If I perform action Ψ, it is very likely that it will be the case that Φ.
- Therefore, I should perform action Ψ.

This year, the project has undertaken a focused program of basic research on the functional components of rational agency. This research is continuous with a large body of previous work done at SRI under AFOSR sponsorship. In particular, much of it connects directly with Dr. Moore's work on the dependence of action on knowledge [8], [6]. The objectives of the research are the development of a richer and more systematic understanding of the roles in the production of rational action of various mental states, such as believing and desiring, and the development of formalisms adequate for representing and reasoning about the concepts involved in the theoretical analysis of rational action.

2 STATUS OF THE RESEARCH EFFORT

2.1 Previous Results

2.1.1 Possible-Worlds Semantics for Autoepistemic Logic

In our previous work [9], we developed a nonmonotonic logic for modeling the beliefs of ideally rational agents who reflect on their own beliefs. We called this system "autoepistemic logic." Defining a simple and intuitive semantics for autoepistemic logic, we were able to show that the logic was both sound and complete with respect to the semantics. However, the nonconstructive character of both the logic and its semantics made it difficult to prove the existence of sets of beliefs satisfying all the constraints of autoepistemic logic. In an effort to overcome the problem, we recently developed an alternative possible-world semantics that enables us to construct finite models for autoepistemic theories, as well as to demonstrate the existence of sound and complete autoepistemic theories that are based on given sets of premises.

The language of autoepistemic logic is that of ordinary propositional logic, augmented by a modal operator L. Formulas of the form LP are interpreted informally to mean "P is believed" or "I believe P." For example, $P \rightarrow LP$ could be interpreted as saying "If P is true, I believe that P is true." If a set of formulas is to be interpreted as a representation

of the beliefs of a rational agent, then a formula LP will be true with respect to a certain set of beliefs if and only if P is in the set. That is, the statement "I believe P" is true for a particular agent just in case he, in fact, believes P. In the original semantics for autoepistemic logic, we simply stipulated that this constraint has to be met by models of autoepistemic theories. This had the effect of requiring that specification of a model include a potentially infinite list of all the formulas of the form LP that were to be taken as true. The resulting lack of structure in the models made it extremely difficult to prove results concerning the models of particular autoepistemic theories.

However, it turns out that, for autoepistemic theories representing sets of beliefs satisfying certain stability conditions, we can define models that have much more structure. The principal conditions are that (1) the set of beliefs is closed under ordinary logical consequence, (2) whenever a formula P is believed, it is believed that P is believed, and (3) whenever a formula P is not believed, it is believed that P is not believed. We have been able to show that a set of beliefs satisfying these conditions can be characterized by a set of possible worlds such that a formula is believed if it is true in every world in the set, and a formula of the form LP is true in a particular world if P is true in every world in the set.

The important consequence of this demonstration is that such a set of beliefs can be characterized by a finite set of finite possible worlds whenever the number of atomic formulas in the language is finite. This in turn let's us define finite models under the same conditions, whereas, under our first definition, the models are finite only if the entire set of beliefs is finite.

With finite models, we can explore certain questions that are much harder to investigate with the infinite models of our original approach. For instance, let us consider what beliefs would be justified on the basis of the set of premises $\sim LP \rightarrow Q$, $\sim LQ \rightarrow P$. Informally speaking, these formulas say "If I don't believe P, then Q is true" and "If I don't believe Q, then P is true." Suppose these are an ideally rational agent's only premises. If he does not believe P, he can reflect on the fact that he does not believe P and he will conclude that Q is true. Conversely, if he does not believe Q, he can reflect on that and conclude that P is true. It thus seems that he has grounds for believing P only if he does not believe Q, and vice versa. So there are apparently two possible stable belief states that can be based on these premises. With the possible-world semantics for autoepistemic logic, we can demonstrate such conclusions rigorously be examining all the possible-world models of the premises. The details were presented in paper [7].

2.1.2 Weak Logics of Knowledge and Belief

Beginning with the work of Jaakko Hintikka in the early 1960s [3], a number of attempts have been made to formulate and analyze varying conceptions of knowledge and belief by using the techniques of modal logic. In such research, the relevant notions are symbolized by intensional operators on sentences. Various axioms governing these operators are then proposed. The important methodological point is that one will be able to apply fairly standard techniques and results from the study of modal logic to the analysis and comparative evaluation of such systems. Indeed, most proposed systems have been exact analogues of

one or another standard modal logic; that is, one simply replaces the modal operator for necessity with that for knowledge or belief. In the case of belief one must drop the analogue of the basic modal principle that, if it's necessary that P, then P. There are, after all, false beliefs.

Though we cannot reasonably idealize false beliefs out of existence, any logic of knowledge and/or belief will have to embody some degree of idealization. Still, it has seemed to many that the commitment to fairly standard modal systems has led to some thoroughly inappropriate idealizations. Two distinct dimensions of idealization have been postulated.

All standard modal logics or logics of necessity are extensions of the system called **K**, which is the minimal modal logic. When conceived of as a basis for logics of knowledge and belief, this system yields the result that the subjects or agents in the intended domain of the theory know or believe all classical logical tautologies and, furthermore, know or believe all the classical tautological consequences of anything they know or believe. With respect to the logic of necessity, these results are widely accepted. Surely all tautologies are necessarily true and, just as surely, if something is a logical consequence of a necessary truth, then it is itself a necessary truth. As applied to logics of knowledge and belief, however, standard systems based on **K** seem to many to require an egregiously inappropriate idealization. Unfortunately, a commitment to working within modal logics weaker than **K** involves giving up some, perhaps a substantial amount, of the power of analysis afforded by standard techniques in the theory of modal logics.

The other dimension of idealization has been that of "introspective" (or reflective) competence. How much are our subjects assumed to know or believe about their own knowledge and/or beliefs? Here too there has been a good deal of disagreement. With regard to knowledge, it has centered upon acceptability of the principle that, if one knows that P, then one knows that one knows that P. (The analogous principle in modal logic is that, if it is necessary that P, then it is necessary that it is necessary that P.) With regard to belief, a further locus of controversy has been the negative counterpart of the foregoing principle; namely that, if one doesn't believe that P, then one believes that one doesn't believe that P. The analogous principle realting to necessity is itself controversial.

Under this project, we have explored less drastic idealizations along the dimension of introspective competence. The considerations that justify commitment to the system **K** as a base are, by and large, purely technical or tacitical–the main point being simply a desire to separate problems that are in principle separable. With respect to knowledge, we suggest that one should begin, at least, with no more than the basic system **K**, together with the principle that, if one knows that P, then P. In the case of belief, more drastic deviations from standard systems are proposed. In particular, a new axiom-called **Y**-is suggested. In one formulation, this axiom amounts to the following: if one believes that P, then one doesn't believe that one doesn't believe that P. This formulation brings out an essential feature of the proposed system: as an alternative to idealizing in such a way as to guarantee great scope to veridical introspection, the suggestion is to idealize in a way that guarantees against false introspective beliefs.

Considerations in favor of such an alternative idealization come from a number of

sources; two, in particular, are the Paradox of the Preface and, primarily, Moore's Paradox. The principle underlying the former is that we don't believe that all of our beliefs are true. Indeed, surely it's irrational for us to believe that we are in no way mistaken in our beliefs. We must therefore reject the principle that we believe that, if we believe that P, then P. (The analogous principle with respect to knowledge is obviously correct.) Moore's paradox consists in this: it is odd or self-defeating for someone to assert both P and that he doesn't believe that P. That is, any utterance of a sentence of the form "P; but I don't believe that P" is, in same sense, self-defeating. The moral of Moore's paradox, at least regards the logic of belief, is that we do not believe of any one of our beliefs that we don't believe it. This is precisely the point of the axiom **Y**.

A development of these ideas was presented in [4]. Both the axiom **Y** and the resulting system $\mathbf{K} + \mathbf{Y}$ are characterized in terms of the now standard model-theoretic techniques for modal logic. This yields both soundness and completeness results. It is shown in what ways the formalization is weaker than standard logics of belief. The paper also contains some more general considerations regarding the appropriateness of modal logics of belief, given varying conceptions of the role of beliefs in action.

2.1.3 Plan Synthesis

Part of our work deals with techniques for automatic planning. Previous work in this vein has been highly experimental in nature, the standard methodology being to explore possible techniques by constructing working programs. Because of the emphasis on experimentation, very little has been done to analyze the techniques to determine why they work, when they are applicable, and whether it is possible to generalize them to solve larger classes of problems. Our work provides at least part of the missing analysis and introduces new techniques for plan synthesis.

We have approached the question of automatic planning from a rigorous, mathematical standpoint. Our methodology has been todevelop a mathematical framework in which to study planning problems, to explore this framework for theorems that can be used to constrain the search for a solution, and then to construct planning techniques based on the theorems that were found. By following this methodology, it has been possible to develop techniques (a) that are capable of solving class of problems than had previously been considered, and (b) that are guaranteed to find a solution if one exists. Furthermore, it has been possible to unify many existing ideas in automatic planning, showing how these ideas arise from first principles.

The mathematical framework that has been developed is very much like that of first-order dynamic logic. In this framework, the world may be in any one of a possibly infinite number of states. Performing an action causes the world to jump from one state to another. A planning problem in this framework consists of a description of the initial state, a description of the goal state, and a description of the allowable actions. The problem is to find a sequence of actions that is guaranteed to force the world into a state stisfying the goal description, given that the world may initially be in any one of the staes satisfying the initial-state description. (State descriptions may be incomplet; that is, there may be more

than one state satisfying a give description.)

Formally, a state description is a set of formulae in first-order logic, and a state is a first-order model. Actions are binary relations on states. For planning purposes, though, all that we need to know about an action are its preconditions and its regression operator. The preconditions of an action are a set of formulae defining the the action may be performed. A regression operator for an action is a function mapping formulae to formulae such that the regression of a formula is the weakest condition that must be true before the action is performed in order for the formula to be true afterward. One of the contributions of our work is a language for describing the effects of an action and a way of computing regression operators from action deescriptions in this language. The language is significant in that it combines the generality of the situation calculus [5] with the notational convenience of STRIPS [2]. This allows the frame problem of the situation calculus to be circumvented to the same extent that it can be done in STRIPS.

The planning techniques are based primarily on two observations. The first is that the world changes state only as the result of an action. Therefore, if a formula is false, it will become true only if an action makes it true. The second observation is that a plan must be finite since we would like our goals to be achieved at a definite point in the future. Consequently, there will always be a last point in a plan when a formula becomes true if it becomes true at all. These observations lead us to the following theorem: a formula is true at a point P in a plan if and only if (1) the formula is true in the initial state and remains true until at least point P, or (2) there is as P that causes the formula to become true and the formula remains true thereafter until at least point P. This theorem tells us that, to construct a plan to acheive some goal, either we must introduce an action that makes the goal true or we must prevent the goal from becoming false it is is true initially. From this theorem it is possible to derive a planning technique. The details were presented in [10]. This work was the core of Pednault's thesis, which was awarded in June, 1986.

2.2 Recent Results

2.2.1 The Role of Propositional Objects of Belief in Action

The essence of the common sense theory of rational action is that such actions are motivated and caused by the beliefs, desires, intentions, and other mental states of agents. Moreover, it is precisely in virtue of the *contents* of these states that the actions so caused make sense. The agent does what it does *because* of what it believes and desires. The import of this can be put as follows. The action the agent performed would promote its desires if its beliefs were true. That is, the agent's action would tend to have the effect that the world came to be the way the agent wanted it to be, if the world were the way the agent's beliefs represent it as being. This core idea can be expressed in theform of the following schema:

- α's believing that Φ and desiring that Ψ cause α to A, and
- if it were the case that Φ, the performance of A by α would bring it about that (or make it more likely that) Ψ would be true.

As the schema is ordinarily–indeed, almost universally, understood, "A" is a schematic letter to be replaced by (nonfinite) verb phrases denoting actions and "Φ" and "Ψ" are schematic letters to be replaced by (declarative) sentences expressing propositions. But, this way of understanding the schema conflicts with two central features of common sense psychology. The first is that the psychologies of organisms, at least within a species, are importantly alike. That is, members of a given species will be caused to act in roughly similar ways, given that they are in roughly similar mental states–given, that is, that they have similar beliefs and desires. The second feature is what might be called a presumption of good design. There is a presumption, that is, that those actions which organisms are caused by their beliefs and desires to perform are, on the whole and in normal circumstances, appropriate, given the organisms' desires and beliefs. Putting the point the other way around, organisms are not built ('designed') in such a way that they are caused, by their beliefs and desires, to perform actions which have little chance of satsifying their desires even when their beliefs are true. These two principles are also central to any effort aimed at the design of intelligent machines, such as (autonomous) robots. Thus, any attempt to design intelligent machines, such as robots, that is guided by the standard misunderstanding of the schema is likely to go astray. Work in this area must be based on a more secure understanding of the role of representational mental states in the generation of appropriate behavior.

The approach we adopt is to begin with simple, partly imaginary, organisms for which all talk of belief and desires can be replaced by talk of perceptual states and of (biologically determined) needs, respectively. Moreover, we attempt to embed this account in a wider, more general theory of information-bearing states and thus to connect work on the theory of action with research in the theory of meaning. Details can be found in the paper "The Role of Propositional Objects of Belief in Action", included as Appendix A.

2.2.2 Events, Situations, and Adverbs

A dispute has arisen over the last few year about the relationship of sentences to the events they describe, and how that relationship is manifested in sentences with adverbial modifiers. The two sides to the argument are what might be called the "Davidsonian position" and the "situation semantics position;" the former being chiefly represented in [1] and the latter by [11].

To summarize the argument, Davidson and Perry agree that sentences describe events, but they disagree as to *how* sentences describe events. Davidson considers and rejects the view that a sentence as a whole describes an event. Instead, he proposes that sentences describing events do so by means of hidden quantification over events, and using this device he proposes an attractive analysis of certain types of adverbial modification. Perry, on the other hand, asserts that the idea that whole sentences describe events is "the only reasonable answer," and—having pointed out elsewhere with Jon Barwise what they take to be the flaws in Davidson's argument—he goes on to provide an analysis of adverbial modification consistent with the holistic view.

One of the things that Perry's and Davidson's analyses have in common is that neither

is based on a very extensive survey of the semantic phenomena to be accounted for by a theory of adverbial modification. The strongest considerations in motivating their analyses are more general logical and metaphysical concerns. A careful examination of the relevant examples, however, shows that neither Davidson nor Perry have the story quite right, and that a more complete account of adverbial modification has to make room for at least two views of the relation between sentences of events, one close to Davidson's and the other close to Perry's. We suggest what such an account might look like. Details can be found in the paper "Events, Situations, and Adverbs", included as Appendix B.

References

[1] D. Davidson. The Logical Form of Action Sentences. In *Essays on Actions and Events*, chapter 6, pages 105–148, Clarendon Press, Oxford, England, 1980.

[2] R. E. Fikes and N. J. Nilsson. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. *Artificial Intelligence*, 2:189–208, 1971.

[3] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, 1962.

[4] D. J. Israel. *A Weak Logic of Knowledge and Belief.* Technical Note 359, Artificial Intelligence Center, SRI International, Menlo Park, CA, August 1985.

[5] J. McCarthy and P. Hayes. *Some Philosophical Problems from the Standpoint of Artificial Intelligence*, pages 463–502. Edinburgh University Press, Edinburgh, Scotland, 1969.

[6] R. C. Moore. *A Formal Theory of Knowledge and Action*, chapter 9, pages 319–358. Ablex Publishing Co., Norwood, New Jersey, 1985.

[7] R. C. Moore. Possible-World Semantics for Autoepistemic Logic. In *Proceedings of Workshop on Non-Monotonic Reasoning*, New Paltz, NY, October 1984.

[8] R. C. Moore. *Reasoning about Knowledge and Action.* Technical Note 191, Artificial Intelligence Center, SRI International, Palo Alto, October 1980.

[9] R. C. Moore. Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence*, 25(1):75–94, January 1985.

[10] E. P. D. Pednault. *Preliminary Report on a Theory of Plan Synthesis.* Technical Report 358, Artificial Intelligence Center, SRI International, Menlo Park, California, August 1985.

[11] J. R. Perry. Situations in Action. unpublished ms. of a lecture presented at the annual meeting of the Pacific Division of the American Philosophical Society, March, 1983.

3 PUBLICATIONS

David J. Israel, "AI Knowledge Bases and Databases", in *On Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Techniques*, ed. M.L. Brodie and J. Mylopoulos, pp. 71-75 (Springer-Verlag, NY, 1986).

David J. Israel, "Notes on Inference: A Somewhat Skewed Survey", in *On Knowledge Base Management Systems: Integrating Artificial Intelligence and Database Techniques*, ed. M.L. Brodie and J. Mylopoulos, pp. 97-109 (Springer-Verlag, NY, 1986).

4 CONFERENCE PRESENTATIONS

Robert C. Moore, "Reasoning About Knowledge in Artificial Intelligence," invited address, Conference on Theoretical Aspects of Reasoning About Knowledge, Monterey, California, March 19-22, 1986.

David J. Israel, "Logic and Intelligence (Natural and Artificial)," invited talk, Association for Symbolic Logic Summer Meeting, Stanford, California, July, 1985.

David J. Israel, "Some Aspects of Non-Monotonicity in Reasoning," invited talk at a symposium on Common Sense Reasoning, IEEE Conference on Systems, Man, and Cybernetics, Tucson, Arizona, November, 1985.

David J. Israel, "The Role of Propositional Objects of Belief in Action", invited address, Conference on Minds and Brains, Indiana University of Pennsylvania, Indiana, Pennsylvania May, 1986.

5 PERSONNEL

The research of this project has been carried out by Robert C. Moore (principal investigator) and David Israel. Supervision has been provided by Stanley Rosenschein and C. Raymond Perrault.

APPENDIX A

By David J. Israel

A The Role of Propositional Objects of Belief in Action

A.1 Introduction

What is the commonsense theory of rational action? There is no short and true answer to this question, but there is a fairly short one that is nonetheless helpful. The commonsense theory of rational action is that

1. The behavior of humans and of other fairly complicated organisms (including, at least in principle, artificial organisms) can be correctly explained, and sometimes even predicted, by reference (and sometimes, in practice, *only* by reference) to the beliefs, desires and intentions - in sum, to the mental states - of those organisms.

2. These explanations are *causal* explanations; the organism's being in mental state Φ – comprising various beliefs, desires, and intentions – caused it to do what it did.

3. Moreover believing, desiring, and intending are all mental states with representational or, as we shall also say, propositional content; so, in particular, a belief represents the world or some part of it as having a certain property. That the world, or that part of it, has that property, is the *content* of the belief state.

4. The causal explanations mentioned in (2) also make sense of what the organism did; the organism did what it did because its beliefs, desires, and intentions had the content they had. Roughly speaking, the action performed by the agent on the basis of its beliefs and intentions would promote its desires if its beliefs were true. That is, the agent's intentional action would have the effect that the world would come to be the way the agent's desires represent it as being if, indeed, it were the way its beliefs represent it as being.[1]

Although there is much that is right with this commonsense theory, there is at least one thing that is wrong with or misleading about it. The model of rational action it puts forth can be schematized as follows:

- α's believing that Φ and desiring that Ψ cause α to A, and

[1]In the foregoing, we've pretended that the contents of beliefs, desires, and intentions were propositional. But this is just a pretense. The proper objects of intention should not be thought of as propositional; as for desires, they are a mixed bag. The proper objects of intentions (and of some desires) are types of acts, as in the intention to take a walk. We shall, for the most part, ignore this. This will be made easier because we shall engage in at least one more gross simplification; we shall speak only of beliefs and desires. We shall thus either disregard intentions completely or pretend that they can be reduced to beliefs and desires.

- if it were the case that Φ, the performance of A by α would bring it about that (or make it more likely that) Ψ would be the case.

A few points about this schema are worth noting. The first is just that it is a schema, and that it is schematic, in particular, with respect to agents. We do tend to presume that normal agents in similar mental states will be motivated and caused to act in roughly similar ways. There is much slack in this presumption, of course; still, we do expect there to be a good deal of lawlike regularity in the psychologies of members of a given species–and, at higher levels of abtsraction, even across species. Thus, members of a given species with roughly similar beliefs and desires tend to act in roughly similar ways. A second point is that the schema is explicit about connecting the rationality of actions with their appropriateness or wellfittedness. If we take seriously that part of common sense pychology that holds that beliefs and desires cause actions, then we had best take seriously the claim that, on the whole, we, and other organisms, are well designed–that is, for the most part or in normal conditions, we are caused to do what, in the circumstances of our action, it is appropriate that we do do, given our desires. Putting this point the other way around, organisms are not built ('designed') in such a way that they are caused, by their beliefs and desires, to perform actions which have little chance of satisfying their desires, even when their beliefs are true.

As the above schema is ordinarily–indeed, almost universally–understood, "A" is a schematic letter to be replaced by (nonfinite) verb phrases of action (e.g., "take a walk", "kiss the Blarney Stone," "talk to his manager"). "Φ" and "Ψ" are schematic letters to be replaced by (declarative) sentences that express propositions. That this is the wrong way to understand the schema is the burden of this essay. Indeed, we shall argue that this way of understanding it is in serious conflict with the two points just made–about the schema's intended generality of application and its explicit commitment to a presumption of good design. We should also note that this mistaken understanding has suggested to some that the proper objects of the propositional attitudes (e.g., of belief) are sentences that express the propositions. It is then further argued that, to account for the applicability of the schema in explaining action among non-language-using creatures or prelinguistic human creatures, one must suppose that these sentences belong to an inner, private, mental language, *a language of thought.* By arguing against the mistaken understanding of the schema of rational action, we hope succeed in undermining some of the appeal of the language of thought hypothesis.

A.2 Information-Bearing States

Clouds *mean* rain. Spots of a certain kind *mean* measles. Smoke *means* fire. The number of rings on a cross section of a tree trunk *bears information about* the age of the tree. The color of a hydrangea *bears information* about aspects of the chemical composition of the soil in which it first grew. The length of the column of mercury in a thermometer *bears information* about the temperature of the surrounding air. In all such cases there is a lawlike or nomological regularity connecting one type of situation with another . Instances

of these regularities are cases in which one situation means something or carries information about another: and, of course, in such cases there need be neither minds nor symbols used by minds. Indeed, such cases are fundamental; the fundamental locus of meaning and information content is to be found in the fact of systematic regularities among *types* of states.

Let us look at one of these cases, that of a red hydrangea. There are a number of things to note. The first is that the cases conform to the essence of Wittgenstein's picture theory of meaning. (Bizarrely ill-named doctrine!) That essence is that it is not *things* that have meaning or represent, but rather structured situations *involving* things (Wittgenstein spoke of facts or states-of-affairs) that represent. Thus, it is not the hydrangea that carries information, it is the *hydrangea's being red* (or the fact that it is red). On the side of the represent*ed*, what is represented (in the primary sense) is not the soil in which the flower grew, but a fact about it–to wit, the fact that that soil had a high iron content (or, *its having had a high iron content*). Other facts about the hydrangea, or other aspects of the hydrangea's state, indicate other facts about other aspects of other things, relative to other regularities.

Before returning to the hydrangea, let us introduce another bit of wisdom from semantic theory. **Meaning is relational.** At first thought, it seems natural to conceive of the meaning of a declarative sentence Φ as that condition (or set of conditions) p_Φ under which Φ is true, and to identify the proposition expressed by the sentence with that condition (or set of conditions). But, on second thought, it is obvious that the truth condition of many utterances varies with variations in the circumstances of those utterances. So, for example, the sentence *I am taller than you* cannot be associated with a single truth condition. Rather, we must say that an utterance of that sentence by a speaker α, in circumstances c in which α is addressing a certain person β a certain at time t, is true if and only if α is taller than β at t. So if David Israel says it *now* to Bob Moore, what he says is true if and only if David Israel is taller than Bob Moore on May 31, 1986. If Ronald Reagan said it twenty years ago today to Barry Goldwater, what he said then is true if and only if Ronald Reagan was taller than Barry Goldwater on May 31, 1966. The sentence is not ambiguous; it is not that it is being used by David Israel with a different meaning from the one it had when Reagan used it. It is simply that different *uses* of it have different truth conditions and that *the difference in truth conditions is systematically related to differences in the circumstances of use*. Thus, the meaning of this sentence can be thought of as a relation between speakers, circumstances (in our case, circumstances involving a parameter for addressees), times, and truth conditions of utterances of the sentence. We shall use notational devices from situation semantics to make this systematic dependence between the values of the mentioned parameters and the truth condition apparent:

$$\alpha,\ c,\ \beta_c,\ t\ [\![\ \textit{I am taller than you}\ ]\!]\ \mathbf{p}(\alpha,\ \beta_c,\ t)$$

In the above, "**p**" stands in for the relation of one thing's being taller than another thing, at a given time. The whole is to be read as follows: α's uttering *I am taller than you* at t to β_c in circumstances c, constitutes making a true statement if and only if α is taller

than β_c at t. (Note that we are assuming certain conditions on α's utterance–roughly, that it be intentional.) Let's connect this bit of wisdom with that coming from Wittgenstein. In general, the meaning of a sentence is a relation between situations involving an utterance or use of the sentence, and the truth conditions of such utterances or uses. Various aspects or features of the structured situation are relevant for various cases; hence it makes sense to highlight those aspects as parameters of the relation, as we did above. Before returning to the garden, note that, if Ray Perrault had just now said, ***He's taller than he is***, looking at David Israel while he said the first ***he*** and at Bob Moore while he said the second ***he***, then what he (would have) said was just what David Israel said. That is, though the sentences we use are not synonymous, what we say, using those sentences, is the same. That is, we both represent the world as being a certain way–the same way. Let us speak of such uses as having the same ***content***, though the sentences differ in ***meaning***. In the examples involving Ronald Reagan and David Israel, we had two uses of the same sentence (with a fixed ***meaning***) being associated with different ***contents***.

Now back to our flowers. If there are two red hydrangeas, h_1 and h_2, side by side in a garden, h_1's being red carries the information about the soil of *its* youth, $soil_1$, that it was iron-laden and, likewise, h_2's being red carries similar information about the soil in which *it* first grew, $soil_2$, i.e, that *it* was similarly laden with iron. Just as before, we have a relation between a type of structured situation, in this case a type of situation involving a hydrangea's being a certain color, and some fact about the world involving parameters systematically related to aspects (parameters) of that situation type. Let us mark the relevant state of the hydrangea in this case as that of BEING RED. So we have the following relation:

$$h,\ l,\ c\ [\![\ \text{BEING RED}\ ]\!]\ \mathbf{p}(soil_h,\ t'_h)$$

Here "**p**" stands for the property of being rich in iron, t'_h is a function mapping a hydrangea onto the time of its youth, and $soil_h$ is a function mapping a hydrangea h onto the patch of soil in which h first grew (which, of course, need not be the soil in which it is now situated). Thus: h's being red at l in c means that $soil_h$, the soil in which h first grew was rich in iron during h's youth. So the state of being red is a meaningful (information carrying) state in hydrangeas; its meaning is a relation between a hydrangea's being in that state and a fact about the soil in which that hydrangea grew. The relation in question is one of causal connection; indeed this example of the relational nature of meaning is an instance of the incontrovertibly relational nature of causation. Moreover, it exemplifies two other aspects of causal relations. One is just Wittgenstein's point "writ large": causal relations hold among complex, structured states or events involving things; it does not hold among the things themselves. The other aspect is that causal relations hold among particular states or events by virtue of the fact that those states or events are instances of certain ***types***.

There is yet a third aspect of the relational nature of meaning that also derives from the nature of causal relations generally. Causal regularities typically operate only against a ***background*** of conditions and other regularities. These conditions and regularities constitute

the normal conditions of the environments within which the regularities hold. (Only the fundamental laws of physics are free from local constraints on their applicability.) Unsupported bodies that are heavier than air fall toward the Earth, but not in spaceships and not near the Moon's surface. In various weird environments, a hydrangea's being red might not be correlated in a lawlike way with the fact that its nurturing soil had been heavy in iron content.

We have noted a number of dimensions along which meaning is relative. Or, to put it another way, we have noted a number of parameters relevant to the information carried by [or relevant to the content of] an instance of a meaningful state. Content is relative both to various aspects of the *circumstances* in which the state is evinced and to various *constraints* operating within those circumstances. Thus, we can speak of the *constraint-relativity* and *circumstantial-relativity* of meaning or information. (When we pass to a discussion of *control states*, we will need to note a third relativity: to function or purpose.)

These two relativities demonstrate that there is something significantly wrong with any view that holds that what the content of an information carrying or representational state is, is a simple matter of fact. Of course, there is also something (arguably the most important part) that is right about it. What's right about the claim is that the determination of the content of a state is not a matter of whim or fancy; to put it baldly, what's right is that there can be a *natural science* of content. What's wrong about it is just that it seems to ignore the relational nature of meaning. That a hydrangea's being red is meaningful at all depends on the existence of regularities connecting the color of hydrangeas with other aspects or features of (perhaps other) things. That a particular hydrangea's being red has its particular content depends both on the particular regularity connecting color and chemical composition and on particlular facts about where that hydrangea first grew. Thus, that h_1 's being red carries the information it does, namely, that $soil_{h_1}$ was, at t_{h_1}, rich in iron, depends on the circumstances of h_1's youth. The constraint, together with the circumstances, yields information about a particular patch of soil, perhaps now widely scattered–information to the effect that it was rich in iron.

We've said that the fundamental locus of meaning and content is to be found in the causal order at large and that, in that order neither humans nor human minds or human languages have any special status. But we have *not* said that this notion of *natural* meaning is all there is to meaning, nor is it all there is to mind.

A.3 Control States

The time has come to establish a connection with commonsense psychology. Hydrangeas, trees, thermometers, clouds, and measles spots are not - let us assume - active agents. That is, they do not engage in intentional behavior. We may sometimes speak loosely and explain their sundry doings by reference to beliefs and desires, but such talk is mere talk. Such entities do not have psychologies. It is otherwise both with people and with other organisms. Let us then add something to our claim that there are meaningful, information-carrying, *content*ful states. As far as commonsense psychology is concerned, it is crucial that such states cause and, at the same time, *rationalize* behavior. We shall speak of mental states

insofar as they function in this capacity as *control states*. As Fodor says:

> *It is characteristic of commonsense belief/desire psychology - and hence of any explicit theory that vindicates commonsense belief/desire psychology - that it attributes contents and causal powers* to the very same mental things that it takes to be semantically evaluable. *[2].*

It makes sense to look first at perceptual states of living organisms before moving on to anything more sophisticated. Such states *are* control states, that is, they are representational states of organisms that are also [partial] causal determinants of behavior. Thus, a certain pattern of irradiation on a frog's eyes *means*, in a certain range of environments, that there's a fly in the vicinity–specifically, in a certain direction from the frog's head. This ocular pattern triggers (causes) an orientation-and-attack response–the frog moves its head in the right direction and flicks out its tongue. The pattern also carries information about the distance of the fly from the frog, but this information plays no role in controlling the frog's behavior. Only orientation or direction counts. (The frog has very little binocularity and hence displays little stereopsis.) So the state carries information about both distance and orientation, for its type is involved in regularities involving both, but only the regularity involving orientation is involved in that aspect of the meaning that controls behavior. We thus have a more complex example of the *constraint-relativity* of content. As before, the perceptual state comprises many aspects or parts. These enter into various regularities and, of course, a single aspect may enter into more than one. Moreover, from the point of view of the control of behavior, *which is the crucial point of view in the case of agents*, not all regularities are created equal. In the foregoing cases, we had no principle for discriminating among regularities. Now we do.

We've talked of the type of state *meaning* that there's a fly in the vicinity. Others have said that what "fly" means to the frog is just that characteristic pattern of ocular irradiation - i.e., as of a small black moving dot. This is just backwards. The facts are that, in a wide range of environments, flies in flight are what actually cause that pattern on the frog's eyes and that *flies on the fly are what the frog is after*. This convergence of the "backward looking" (environment-caused) and "forward looking" (behavior-causing) aspects of the state is a good thing (from the frog's parochial point of view, of course). Moreover, it's a good thing that (typically) flies do not fly in slightly dispersed two and threes; when, for example, two such potential tidbits are buzzing around, the poor frog strikes out in a direction midway betwixt the two. On the other hand, it's nice that flies sometimes congregate thickly enough so that the frog, when confronted by a swarm, can strike out with a good chance of dining success. So in these regards, at least, Nature has been kind to the frogs.

Now these latter considerations introduce a third and quite different dimension of relativity–*relativity to purpose*. Why should the frog be able to respond differentially and appropriately to the presence of live flies? Well, because it needs to eat and the flies are food. Why does it need to eat? Ultimately, the frog needs to eat to stay alive. Why does it need to stay alive? If you like, it only really needs to eat to stay alive long enough to

make its contribution to the gene pool. (Of course, the frog doesn't care either a fig or a fly about *that*. It's just that well-designed frogs tend to survive long enough to reproduce.) This, or something like it, is the end of the story. We shall, in what follows, relate this talk of purpose to talk of desire and, indeed, talk of the latter will replace talk of the former. This is done simply to sidestep a very big issue, namely, that the primary notion here is precisely that of pupose, primary with respect to *both* belief and desire. After all, why do organisms have belief-desire states at all? And why do they have the belief states and desire states they do? In any event, the essential point is that what we are after here are generalizations linking beliefs, desires and *appropriate* behaviors – appropriate, that is, for promoting the desires and realizing the needs of the organism. We shall simply note here that the meanings we shall attribute even to the backward-looking (perceptual) aspects of the frog's control state are themselves theoretically controlled by (i.e., are relative to) the problem that the frog's visual system must solve in the hypothesized environments. In other words, the attribution of meanings to such states is conditioned by the purposes that the visual system subserves.

Now, let's imagine two frogs, Kermit and Prince, in different spatiotemporal locations l_1, l_2, in different, though equally normal circumstances, c_1, c_2. Kermit's circumstances contain a lone fly, Henry, and Prince's a lone fly, Dick. Henry is flitting around near Kermit, Dick, near Prince. Henry's flitting about has caused a particular instance of the relevant state–let us call the type Φ–in Kermit; and likewise with the Prince-Dick duo. The instance of the state type in Kermit at l_1 in c_1 carries the information that a certain fly–to wit, Henry–is flying around to his (Kermit's) right at l_1. The instance of Φ in Prince, at l_2 in c_2, has the content that a certain fly, Dick, is flying around to his (Prince's) right at l_2. Kermit's state causes a particular bodily movement (of Kermit's body, of course), a movement of his head to the right and a quick protusion of his tongue. Prince's state causes an instance of the same type of bodily movement, that is, it causes a movement of *his* head to the right and a quick protusion of *his* tongue. Let us call the type of bodily movement Ψ.

Regularities lurk: regularities involving Φ, regularities involving Ψ, and crucially, a regularity linking the two–that is, connecting perception and movement.[2] If we want to identify and understand that regularity, it must be at the level of *meaning*, not at the level of propositional content or information carried. Thus, Kermit's and Prince's respective states do not carry the same information; they do not have the same content. One carries information about Henry's flying about at l_1, the other about Dick's flying about at l_2. But they can be said to have the same *meaning*: each means that there is a fly flying about nearby and in a rightward direction.

Nearby in relation to what or whom? In a rightward direction relative to what? Just as with the meaning of the English sentences in the examples above, to identify the relevant regularity, we shall treat the meaning of the state *relationally*. The parameters of the relation must include (at least) a subject of visual experience, a spatiotemporal location,

[2]Where is the desire, you might ask? For the moment, we shall simply assume that a frog is always hungry enough to go after the stray fly. This is actually close to the truth: it is important to note, in this regard, that the bodily movement in question costs the frog very little, in both energy and time.

and an object in the circumstances whose movement causes the pattern. If Nature is kind (to the frog at least), *that object will be a fly*. In any event, that object is also an aspect of the circumstances and a constituent of the content carried by a production of Φ in those circumstances. Thus we obtain

$$\alpha, l, o_c \; [[\; \Phi \;]] \; \mathbf{p}(o_c, \alpha, l)$$

In the above, "**p**" stands for the relation of one thing's flying about to the right of another at such and such a spatiotemporal location. One could have abstracted direction relative to α as a separate parameter as well. We would then be dealing with a different state–not one that meant, for a given α, that an object was to the right of α, but rather one that meant, for a given α, that an object was at some direction δ relative to α. Things are complex enough as it is. (For a bit more along these lines, see below.)

Φ is a structured st[illegible] **type**; in particular, it is a type of structure manifested in frog's brains–well, actually on the surface of their eyes. It is not a structure whose every manifestation (token, production) bears the same content; it is rather a structure the content of whose productions is determined in crucial respects by the circumstances of those productions. In this respect, therefore, it is like all sentences of every natural language. (Now there's a linguistic universal for you!!)

We've already remarked that what *we theorists* need to get at the relevant regularities are precisely such meaningful types–types whose meaning is a relation between circumstances and contents. **But such states are also what the frog needs.**

A.4 Meaningful State Types Versus Propositions

If one identifies propositions as the contents of instances of meaningful states (hence, in particular, as the truth conditions associated with statement-making utterances of sentences), then it's quite clear that propositions are useless by themselves for guiding or controlling behavior. Imagine some mental state whose meaning can be identified with its content–that is, whose meaning is a constant function from arbitrary circumstances to a given proposition. (Such a state is the analogue of a so-called *eternal sentence*; this is a sentence all possible utterances of which, no matter what the circumstances, express the same proposition–i.e., have the same information content.)

Let's fix a particular content. We are to imagine a state type, call it Φ', whose content is *always, in all circumstances*, that Henry is buzzing about at such and such a spatotemporal location. The location, of course, must be given in terms of some non-agent-centered (in particular, non-Kermit-centered) coordinate scheme, say in terms of seconds of longitude and latitude, and years, days, hours, minutes, seconds of Greenwich mean time; the "is" in the above specification is in the infamous "timeless present" tense. So the conmtent of Φ' can be thought of as identical with the actual content borne by the instance of Φ in c_1 at l_1.

Now, how can Kermit's believing *this* proposition in *this* way guide his behavior; that is, how can his being in a state of the kind described be of any use to him? (We are *not* yet

asking how being in a mental state with that content as its meaning could possibly guide the behavior of *any* entity. For the moment, we're sticking to frogs.) It adds naught, of course, if we add a non-Kermit-centered scheme relating the relative orientations of Kermit's head and Henry. **Nor does it do any good to assume Kermit has in his little head a token of an eternal sentence with that content.** States containing such tokens are, by themsleves, useless and nothing short of magical powers in the interpreter processing those tokens will convey utility on them. Remember that, by hypothesis, the particular instance of perceptual state Φ that Kermit is in, under the circumstances at the spatiotemporal location described, has precisely this content: a certain fly, Henry, is flying around at, say, 10:30 a.m. Eastern Standard Time on May 31, 1986, at such and such a longitude and latitude (down to the millisecond, no doubt) and in such and such a direction (relative to what–the fixed stars?). What his being in that state then causes him to do constitutes appropriate behavior for him in those circumstances and in view of his hunger. But are there any bodily movements that are appropriate to an arbitrary and arbitarily located fly-deprived froggy agent by virtue of its being in a state that carries *that* information, that is, by virtue of its being in a state of type Φ'? Or, to put it more commonsensically, is there anything interesting to be said about what a hungry frog that believes *that* proposition *should* do? Obviously not. Just so: there is no single thing or interestingly uniform collection of things to be done by everyone who not only has David Israel's interests at heart, but also believes that David Israel is in EK268 at SRI International in Menlo Park, California, on May 31, at 1:45 p.m. PDT and, moreover, has the additional belief that a bomb is set to go off in EK268 at SRI International in Menlo Park, California, at 1:50 p.m. PDT, on May 31, 1986. Quantifying over propositions or belief states characterized in terms of their *contents* is useless. Actions are not to be explained by the em propositional contents of beliefs and desires, nor are they predictable on that basis. They are rather to be explained (and predicted) in terms of the meanings of meaningful physical states of organisms.

What Kermit needs is a control state whose *meaning* can be characterized roughly as follows: there's a fly, off to the right (or, there's food off to the right.) That is, he needs a state, instances of which will be triggered whenever (and perhaps only when) there is a fly off to his right and instances of which will reliably produce appropriately fly-catching bodily movements. Such a state he has: Φ.

A.5 Content Without Representational Aspect

Note that neither we theorists nor Kermit needs a state whose meaning is to be characterized as follows: there's a fly off to *my* right (or, to the right of *me*.) There is no need, that is, for the state to be structured in such a way as to contain a component representing the agent whose state it is, either parametrically or absolutely (nonparametrically). Thus, the instance of Φ in Kermit need contain neither a part representing himself, nor a part playing the role of a parameter for the subject of the visual experience. So far, at least, Kermit has no need for a concept of himself; ***it is quite enough that he is the one in the state.*** Furthermore, there is no need for the state to have a component representing the relevant spatiotemporal location. This location, which is the one the information carried is about

and at which Kermit's bodily movement (and action) takes place, is determined simply by the facts and the relevant regularities. To quote the Bard: What is l_1 to Kermit, or Kermit to l_1 - *except, of course, for being where he is at*? For the same reason, there is no need for the structure of the state to contain a parameter for spatiotemporal location. There is no need (so far, at least) for Kermit to have "beliefs" about where he is–not even *indexical* beliefs to the effect that "I am at such and such a place." Indeed, we do not have to suppose that there is any aspect of his state that we might characterize as functioning like the indexical "here".

A similar point can be made about the relation of one thing's being or flying around to the right of another. In this case, the state does have an aspect correlated with direction–in particular, with BEING TO THE RIGHT. But of course, being to the right is not an intrinsic property of things; indeed, the relation in question seems irreducibly three-placed; x is to the right of y from z's perspective. The cases we have in mind are those in which y and z are identified and, moreover, are anchored to the subject parameter. In our case, it is Kermit's own perspective, with his head pointed as it is, that determines what is to *the* right. So here we have a three-place relation being represented by a state containing an aspect correlated with a single parameter of the relation–the object and cause of the experience. Actually, it's easy enough to make a case for a fourth parameter of the relation–the orientation of z; after all, z might be hanging upside down. This last parameter can be taken as fixed; at least, it can be so taken in any of a wide range of environments–in particular, environments sufficiently within the earth's gravitational field to ensure a fixed rightside-up orientation relative to the earth's center of mass. Of course, we are also supposing that frogs, unlike opossums, are mostly upright creatures. What fixes this parameter, then, is an aspect of the background against which the generalization that subsumes our frogs' behavior holds. Given that it is so fixed, it seems needless profligacy to assume that an aspect of the state must correspond to it. To handle frogs in space, *we theorists and designers* would have to make this fourth parameter explicit. In sum, therefore, we can say that the mental states of subjects need not have components corresponding to every component of the content carried by instances of those states–indeed, not even to many of them. (For more on such points, see [3], [5], [6].)

A.6 Types of Bodily Movement vs. Actions

We've said that what both we theorists/designers *and* Kermit and Prince need are *meaningful* mental states–structures whose instances have contents determined in systematic ways by aspects of the circumstances in which those instances are produced. The same goes for the "forward looking" aspect of the state, the role the state plays in triggering bodily movement. Being in a state of type Φ causes a certain bit of bodily movement, but the bodily movements particular instances cause are different. Kermit's being in state Φ at l_1 causes his body to move in a certain way at $l_{1'}$. Prince's state at l_2 causes his to move in a certain way at $l_{2'}$. Here too we have relativity to context and constraints. Let us call the type of bodily movement in question, that of turning one's head to the right and sticking out one's tongue, Ψ. (We can, if we like, assume that the specification of this type

is species-specific.) Just as we conceived of the state Φ in its information-bearing aspects as relational, so too can we think relationally of the type of bodily movement typically caused by being in Φ. Types of bodily movement are relations among an agent α, circumstances c, spatiotemporal locations l_c, and actions performed. One can even unify a little here by having the last argument be the proposition that such and such an action is performed. Let's look at the case in hand. We have:

$$\alpha, l_c, c \, [\![\, \Psi \,]\!] \, \mathbf{r}(\alpha, o_c, l_c)$$

We shall fix Kermit as our agent and $l_{1'}$ as our location; the circumstances are as described above, with Henry flitting about to Kermit's right. In this case, if all goes well, the bodily movement of type Ψ results in or constitutes Kermit's catching Henry with his tongue; that performed by Prince in his circumstances constitutes or results in *his* catching Dick with *his* tongue. (Or we can say that the bodily movement determines as true the proposition that Kermit catches Henry with his tongue at $l_{1'}$, etc.) Given that this is the kind of *action* involved, we do well to make explicit the parameter for the object of the action–in our case, poor Henry, but, in general, whatever it is that ends up on the frog's tongue when the action is successful. Note that the object is not a parameter of the bodily movement Ψ; it does not correspond to an aspect or part of Ψ. It is, however, a parameter of the action performed; it is that thing caught on the agent's tongue. The action is that of catching *something* on (or with) one's tongue.

Let's step back from the particular to the general. What is the meaning of a type of bodily movement? An odd question to ask, you may think. But not so in computer science. The idea of treating the meaning of a program or more generally, of a programming-language construct, as a relation among *computational (machine) states* is central to theoretical work in computer science. The crux of this idea is precisely that the type of activity consisting of the execution of a program or instruction can be represented as such a relation. Research in artificial intelligence on planning has developed a similar idea: that of act types as relations among states of the world. The version of this general idea introduced here, that of treating the meaning of a type of bodily movement as a relation between circumstances and resulting action (or a proposition to the effect that such and such has been achieved), was first proposed by John Perry.

If we theorists are to explain the success (such as it is) of frogs, all we need do is compose relations, relating various parameters of the composed relations. Thus, the temporal location of the bodily movement must be fixed to be just after the ocular irradiation (actually about 100 milliseconds afterwards.) Moreover–and this is especially crucial–the object parameter of the *action* must be coassigned to the parameter for the object in the circumstances which causes the irradiation–which parameter, in turn, must be coassigned to the parameter involved in the propositional content carried by the particular occurrence of Φ. The object caught on Kermit's tongue is the very same as the object (Henry) whose movements caused Kermit to be in a perceptual state of type Φ. That state, in turn, caused the bodily movement of type Ψ that constituted, in the circumstances of that movement, Kermit's catching that object (poor Henry again). In the circumstances in question, Kermit's

producing a bodily movement of type Ψ makes true the proposition that Kermit catches Henry with his tongue at l_1, (and is thus enabled to commence dining.)

This, of course, is just what Kermit *wanted* or *desired* or even *intended* to accomplish: he desired to catch and eat that fly. Or, to resume our fiction with respect to the proper objects of desire: Kermit desired that he (Kermit) catch and eat Henry. Let us now pay a bit more attention to the appetitive aspect of things. Since we assume that Kermit is hungry, we can assume an appetitive state, call it Ω, whose approximate meaning is (in line with our fiction) **that I eat a fly**, or, on the assumption that live flies are the frog's sole or major food, **that I eat**. Imagine Kermit in such a state at l_1; this is to imagine Kermit's mental state at l_1 as comprising both a perceptual and an appetitive aspect. Kermit's being in such a state, in those circumstances, causes him to do what he does, that is, it causes an execution of Ψ. Now just as we speak of the perceptual aspect of the total state, as produced in those circumstances, as carrying information about Henry, so too can we speak of the appetitive aspect of the state, in those circumstances, as being *directed toward* Henry. That is, the desired state of affairs is one in which Kermit catches Henry on his tongue and eats him. This is no case of mere notional desire. Kermit doesn't desire merely that he eat a fly, some fly or other–though this is the *meaning* of his appetitive state. In the circumstances depicted (and assuming that we are to use desire talk at all here) we can say that there is a particular fly Kermit desires to eat; that particular fly is the one he sees, and that fly is Henry. There is a fly, namely, Henry, the fly he sees, such that Kermit wants to eat *that fly*.

A.7 Putting the Story Together

The whole story, up to this point, might be put as follows.

If things go well for the hungry frogs of this world:

> For any frog α, at any time l_c, in a wide range of normal circumstances c, the typical cause of a certain pattern of ocular irradiation Φ is a living fly o buzzing around to the right of α (to the right of α from α's own perspective). Moreover, given that the frog is hungry–i.e., given that it is also in state Ω–being in state Φ very quickly causes a characteristic type of bodily movement Ψ, of α's moving its head to the right and shooting out its tongue, which in c results in α's catching o on its tongue.

For future reference, let us note that the backward looking aspect and the forward looking aspect meet up at o. The cause of the perceptual state is the [unlucky] object of the successful action.

The crucial point in the foregoing is that the generalization cannot be stated at the propositional or content levels; the crucial generalization links state types, construed relationally, and types of bodily movement, also construed relationally. Being in such and such a type of state causes such and such a type of bodily movement: what content being in that state carries depends on the circumstances in which one is in a state of that type;

what action one accomplishes *by means of* the bodily movement (or *in* moving that way) depends on the [related] circumstances in which one executes a bodily movement of that type. But the generalization has a purely *psychological* or internal aspect; being in a mental state which is of types Φ and Ω causes a bodily movement of type Ψ. We can, if you like, make it even more pristinely psychological by characterizing the effect in terms of firings of such and such efferent pathways. Flies, spatiotemporal locations, and the external world as a whole, do not enter into the story *at this level.*

Fodor makes the following comment:

> An explicit psychology that vindicates commonsense belief/desire explanations must permit the assignment of content to causally efficacious mental states, and must recognize behavioral explanations in which covering generalizations refer to (quantify over) the contents of the mental states that they subsume. [2]

We can now see that Fodor is right in the first part and unclear in the second. If, by content, Fodor means anything like information carried or proposition determined to be true by the occurrence of a meaningful state, then the second part is simply wrong. Our generalizations do not involve quantification over or reference to propositions. Rather we refer to state types or, if you prefer, to the *meanings* of meaningful states.

A.8 The Schema Revisited and Revised

Let us return to the schema of rational action. We can now see how to repair the schema in such a way as to capture the relativity to circumstances of informational and control states, as well as action.

- α's being in a mental state which, in circumstances c_1 constitutes α's believing that Φ and desiring that Ω causes α to engage in a bodily movement Ψ, which, in circumstances c_2, constitutes α's A'ing, and
- if it were the case that Φ in c_1, α's A'ing in c_2 would bring it about that (or make it more likely that) Ω would be true.

A.9 Some Corollaries

There is, believe it or not, a great deal yet to be said about even this simple case. Let me highlight just a few of the more noteworthy points. The state Φ is obviously a complex, structured state. As we have seen, not all aspects of its structure play a role in behavioral generalizations in which the state type figures. Obviously, some aspects play no role at all therein. On the other hand, certain aspects do: The pattern as of a small, moving black dot is one such; another is the direction in which the dot is moving (over the very short span the frog requires). The latter controls the crucial parameter of the bodily movement, i.e., the direction in which the frog moves its head. The former might be said to trigger the movement. Now any other fly, in those circumstances, moving in that way (at roughly that

distance from the frog), would have caused *an* instance of Φ, not the instance we've been discussing, but an instance, perhaps qualitatively indistinguishable from the actual instance along all dimensions relevant to the control of behavior. That instance, so caused, would have had *that* fly as a constituent of its content. Still, it is Henry that is in fact the object of Kermit's vision and desire. It is Henry who caused that particular instance of Φ, not another fly. It is at Henry that Kermit's bodily movement is directed. It is Henry too, who, if all goes well (though not for Henry), gets caught and eaten, and thus it is Henry who will satisy Kermit's longing. This is simply the point made above: the "backward looking" and "forward looking" aspects of Kermit's cognitive state meet, in the happy (for Kermit) confluence of circumstances, at Henry. Given all this, perhaps we can speak, in such a case, of that aspect of the ocular event as *standing for* or *denoting* Henry. (Here we have perhaps uncovered the real roots of reference.)

Once talk of the relation of *denotation* has been introduced, can talk of *compositionality* be far behind? Indeed, one could go into an account of compositionality even here. To repeat: the perceptual state Φ is a complex, structured state. Various aspects of that state are correlated with various parameters; asoects of particular instances of Φ are then connected, by way of the attendant circumstances, with particular values of those parameters–in our example, Henry with o_c. In different circumstances, other things are determined as values of those same parameters–Dick, in the circumstances surrounding Prince, and perhaps Dick again in a counterfactual circumstance involving Kermit. The *content* of α's being in state Φ in c is determined by facts about aspects of Φ as well as about the relation between those aspects and aspects of the circumstances in which the instance is produced. Now these aspects can be thought of as parts, in a sufficiently broad sense. Thus, the content of an instance of the whole state Φ is determined by the contents of some of its parts. That is the essence of compositionality.

A.10 The Problems of Falsity and Error

So far, we have been assuming that nature is very kind–to frogs, at least. The problem is that frogs are sometimes led to "believe" that there are flies in their neighborhood when, in actuality, there are none, or that there is food available in a certain direction when there is not. As soon as we have cognitive states with both backward and forward looking aspects, there arises the possibility of *significantly* false beliefs–for in such cases, there arises the possibility of *inappropriate actions*. Note: falsity can arise even in the cases of tree rings and hydrangeas, but in those case, it is sterile falsity–falsity to no behavioral effect. Talk of belief is essentially functional talk: the crucial function (broadly and loosely construed) of belief states is that they represent the world as being a certain way and, together with desire states, cause bodily movements. What movements? Well, what movements should they cause? If things go well, they cause those movements which, if the world is as it is represented, will constitute the performance of an action that satisfies the agent's desires. If the world is not the way it is represented as being, the bodily movement is considerably less likely to succeed.

Given our emphasis on the control of behavior, we need not worry, for instance, about

information regarding the distance of the cause of an instance of Φ from its subject that is carried, in the circumstances, by that instance. That is, in terms of control, the target object is not represented as being at a certain distance from the frog. (Here too, the happy fact is that, often enough, it is a little thing like a fly, rather close to the frog, that causes the characteristic pattern.) So too, even if we are prepared to attribute beliefs to frogs at all, it is in a sense capricious of us to attribute false beliefs about distance to the frog, even in those cases in which it is distance that impedes success, as it sometimes does. But it is otherwise with direction. This is the critical control parameter and, as we noted above, frogs can be "fooled" by scattered pairs and trios. Here we have a real mismatch between the informational, backward looking, aspects and the control or forward looking aspects of the frog's mental states.

Let us imagine the following scenario: Kermit becomes aware of Henry buzzing around to his right. In the 100 milliseconds it takes to activate his tongue, a lot of things are happening outside Kermit's ken. Henry drops immediately to the ground and some other fly, George, comes zipping in from farther afield to Kermit's right and then zigs right. Kermit, by a happy accident, catches George. Here the object in the environment that caused the instance of Φ is not the same as the object involved in the successful action that satisfies Kermit's hunger. Thus, we do not have that convergence at a single object of the content of the proposition believed and the action actually performed that was to comprise the basis of our talk of a primitive mode of denotation. Two different objects are assigned to the two distinct object parameters involved in the two types. Did Kermit get what, in the circumstances, he wanted; did he satisfy the *content* of his desire?

Alas, he does not, but why should Kermit care? Of course he did sate his hunger; look at the smile on his face as he happily gobbles up George. The action actually performed, catching George on his tongue (followed, no doubt, by greedy gobbling) promotes the goal of getting something, indeed a frog, to eat and that is the *meaning* of the desire component of Kermit's mental state. But Kermit was aiming at Henry. The desired action was catching Henry, who, after all, had caused all this activity in the first place. Henry it was who induced the perceptual state that, given the desire for a fly, in turn caused the bodily movement of moving the head and extending the tongue. Henry it was who was involved in the content of the belief, i.e., in the proposition that, by virtue of his being in that state at that time, Kermit believed. Still, Henry did not enter into the *meaning* of that state. No more than Ω (or Ψ) does Φ contain a meaningful aspect whose meaning is a particular individual object (fly). The property of being a fly is involved in characterizing the meaning of the perceptual state. (At least it is so, if one assumes that, in the range of environments being considered, only flies are *both* characteristic triggers for the perceptual state in frogs *and* froggy foodstuff.) But the property of being identical to Henry is not so involved. The same goes for desire: what is desired, i.e., the meaning of the desire state Ω, is to eat (some food, namely, a fly).

A.11 On the Role of Individuals (and the Semantics of Proper Names)

Indeed, what is true for frogs *may* be true throughout the animal kingdom, up (?) to and including us. Individuals are not components of the *meanings* of mental states.

Now let us pause for a brief digression into semantics. Our case of mistaken identity (of George for Henry) is supposed to remind the reader of Donnellan-type examples, in which the semantic referent of a singular term is distinct from the speaker's referent–distinct, that is, from the object the speaker has it in mind to talk about. Well, whether it reminds the reader or not, it does serve to introduce a correction of a widespread misunderstanding about the semantics of proper names. We have said that particular individuals are neither components of nor involved in the meanings of meaningful states. Likewise, particular individuals are not involved in the *meanings* of singular terms of natural languages – not even of proper names. There has been some confusion about this. Kripke has convinced many (and rightly too) that the semantic value of a use of a proper name is an individual. Unfortunately, he seems to have convinced many of these same people that the *meaning* of a proper name (i.e., the semantic value of a name in the language) is also simply an individual. But this last conviction is mistaken; at the level of the language, no particular individual named, say, *David* or even *David Joel Israel* appears. It is not a fact of English that *David Joel Israel* stands for me. After all, as far as the language is concerned, there could be lots of people with that name; again, as far as English is concerned, David Israel might not have been one of them. Perhaps the *meaning* of a proper name β in the language is the property of being named β. At any rate, this isn't a bad first approximation.) Any particular use of the expression β, as a name, has as its semantic value (i.e., refers to) some individual or other. (If things go well, that individual will indeed have the property of being named β, although it need not. Thus, as Kripke points out, Donnellan type examples of various sorts involving proper names are possible too.) The same goes for the aspect of Φ that is correlated with the object parameter. Its meaning might be fixed as follows: for any instance of Φ, it is that object whose movement causes that instance. Its semantic value is not some particular fly (e.g., Henry), not even some particular fly whose movement happens to cause some instance (e.g., Henry).

A.12 What Must an Agent Believe?

Back to the simpler world of frogs and flies and to the case of Henry and George. Ask yourself the following question: What beliefs would Kermit have to have, such that their joint truth would guarantee that the bodily movement executed would, in the circumstances, actually constitute performance of the desired action. (Remember: the action desired is to catch *that* yummy fly, Henry.) Well, Kermit would have to believe that the fly toward which he was extending his tongue was the very same fly that had caused the triggering perceptual state. Maybe Kermit must also believe that that fly (Henry) kept flying to his (Kermit's) right, and that it (Henry) ended up exactly where it should have, following the projected trajectory, 100 milliseconds down the road. Moreover, of course, Kermit would have to believe that that fly was in reach of his tongue, etc. We shall return to this last

bit soon, but first we focus briefly on the attributed belief in the identity of *that* fly, the one Kermit saw, with *this* fly, toward which Kermit aims. **These are propositions that must be true if Kermit's bodily behavior is to constitute his performing the desired action; but are they also propositions that Kermit must believe?**

To see the crucial point more acutely, we might want to do a little fictional ethology, and imagine that frogs track their prey over more extended periods, including times when the prey is invisible–that is, when it is not inducing any perceptual state in the frog. Now let us assume that the tracking case requires that the frog's states over a stretch of tracking be characterized as carrying the information that one and the same prey is being followed. That is, in the extended-tracking scenario, we would need to characterize some aspects of some of Kermit's cognitive states as meaning, roughly, " *This* fly is the same as *that* fly." Such states would require aspects for two related parameters: the present object of visual attention and a previous object of visual attention. Of course, what entity is seen (at a given time) is determined by the circumstances in which the perception takes place, and two distinct things can be seen in two sightings, even when these occurrences are consecutive. Here the phenomenon of memory or information storage rears its interesting head. To develop this tracking story further would require a lot of time and effort; at the end of this report, We shall return to this story as revelatory of a more general phenomenon. Instead, we now to a simpler and somewhat more manageable extension of our story.

So, let's do a little less complicated fictionalizing. Let's endow the frog with a greater degree of binocularity and let us assume that the distance from the frog's head *is* a control parameter, controlling the degree to which the frog stretches out or uncurls its tongue. So, *it seems*, we must now consider another aspect of Φ, together with a new constituent of the content carried. Moreover, we might also want to consider a new dimension along which bodily movements can be discriminated: the extent of the tongue's reach. (But note that, if we do this, we still do not need to add any such dimension at the level of action performed. The *content* of the desire aspect of the mental state to be satisfied is still to catch *that* fly.) Let's try to separate the elements in this compound more clearly.

We imagine that frogs have tongues of varying lengths, or that they differ in the length to which they can extend their tongues. Returning to Kermit and Prince, let us assume that they do indeed differ in that respect, in particular Kermit's tongue is longer than Prince's (or perhaps Kermit can simply stretch it farther). Say Kermit's tongue is 6 inches long, Prince's is 4 inches. Let us arrange their circumstances so that their respective flies, Henry and Dick, equidistant from their heads–that is, the distance between Henry and Kermit's head equals the distance between Dick and Prince's head. Let this distance be 5 inches. If Prince sits still, he's out of luck and Dick is in luck. There is something amiss and it's not Prince's *aim*.

We have two circumstance parameters, c_Φ and c_Ψ, that associated with Φ and that associated with Ψ. These are 100 ms apart, but there is a more important difference between them that the possibility of falsity brings out. The first parameter bears the burden of mediating between the inner state of the subject and the proposition believed (or the way the world is represented as being); the second bears the burden of mediating

between bodily movement and the performance of the desired action (or attainment of the desired state). We assume that Prince's perceptual state carries the information that Dick is 5 inches out to the right, just as Kermit's carries the information that Henry is 5 inches out to *his* right. The circumstance (or aspect thereof) that mediates between the distance aspect of the perceptual state and the distance component of the content carried, is the value of the distance parameter in the circumstances, namely, 5 inches. But the value of that same parameter *required* to mediate between the type of bodily movement *performed by Prince* and *the successful capture* of Dick is 4 inches. (We are assuming that both Kermit and Prince stretch their respective tongues out all the way.) That is, the proposition that would have to be true for Prince's sticking his tongue out all the way to result in Dick's capture is that the ill-fated fly is (no more than) 4 inches away from the hungry frog's head. Unfortunately, for Prince to believe this proposition is for him to indulge in *wishful thinking*. This introduces only a very slight tension with respect to content ascription. From the standpoint of information borne by its perceptual aspect, the state indicates that Henry is 5 inches away; from the standpoint of the prerequisites for success of the bodily movement caused by the state, Henry is (or had better be) no more than 4 inches away. A violation of the law of noncontradiction? Hardly. It is rather a confirmation of the reality of the partiality and the constraint-relativity of meaning and information. The total mental state in question has both perceptual and control aspects. And these aspects are involved in diverse regularities–regularities that, when things go well, merge in a certain way. Still, there's more going on here than that.

If Nature wanted to be perfectly kind to all frogs equally, what would or should she do? One suggestion is that she could adjust things so that the distance aspect of the perceptual state would be in harmony with the distance aspect of the control state. That is, the perceptual states of differently endowed frogs would differ along a certain dimension; things 5 inches away would look a little different to Prince than they would to Kermit. For Prince, they would look a little too far away to go after merely by sticking his tongue out. As a result, he would either not waste his time on the distant Dick, or (and this is a slightly different story) he would compensate by craning his head forward a bit as he moves it to the right. Thus, Nature might not allow Prince and Kermit to be in the same perceptual state in similar circumstances. Furthermore, she might engage in some beneficent "twiddling" with the connections between perceptual states and bodily movements.

That's one thought; here is another. The perceptual states are the same, but, while Kermit and Prince both extend their tongues as far as possible, according to another, more refined system of classification, the type of bodily movements in a frog of Kermit's tongue length is not the same as that in a frog of Prince's tongue length. One type involves the agent's extending his tongue 6 inches; the other involves the agent's extending his tongue 4 inches. (Of course, Kermit can exhibit both types.) So, we could imagine a benevolent Nature having things work out differently in the two cases, without having to alter the perceptual state. Frogs of Prince's type should exhibit a different type of bodily movement when in state Φ (and hungry) from the one executed by frogs of Kermit's ilk; they should crane their necks forward a bit (an inch) and then stick their tongues all the way out. That

is, being in state Φ should cause a bodily movement of type Ψ' in frogs like Prince, where Ψ' consists of turning to the right, craning one's neck forward an inch, and letting one's tongue shoot out as far as possible.

Fine and dandy for Prince and his ilk; a deep issue for us. In our example there was a mismatch between, on the one hand, the circumstances mediating between mental state and the proposition believed and, on the other, the circumstances required to guarantee that the type of bodily movement produced would constitute performance of the "intended " action (would satisfy the desire). In our proposed remedy of this mismatch, we referred to a harmony between aspects of perceptual states and controllable bodily characteristics. As we suggested, this accommodation can be accomplished in more than one way. Still, this sort of thing can only go so far.

How far? Well, remember what is at issue here. We have two types of circumstance, a meaningful cognitive state and a type of bodily movement. We have so parceled things out that one crucial aspect of the cognitive state is functioning rather like belief; its content represents the world as being a certain way. The appetitive aspect of the state, on the other hand, is involved in the specification of the action to be performed to satisfy the appetite; it therefore represents the world as satisfying the agent's desire by virtue of the agent's action. In each case, the circumstances mediate; in the first case, between the belief state and the content of the belief–in the second, between the bodily movement and the action performed. Now, one suggestion might be that the belief aspect should be determined as follows: the content carried by an instance must be such that its truth would guarantee that the bodily movement executed will, in the given circumstances, constitute the performance of the desired action.

John Perry, in [4] considers (though only to reject) the suggestion we're now examining.

> The job of closing the gap between the behavior a cognitive state causes and the goal it is to promote, should be borne mainly by the proposition believed. That is, the truth of the proposition believed should guarantee that the agent is in those circumstances in which the behavior caused promotes the goal desired.

The first of the two above-mentioned ways of accommodation of the mental lives of frogs to differences in their tongue lengths can be seen as a simple instance of this strategy. That is, we can imagine that, in the circumstances envisaged, Prince believes that Dick is too far away or even that Dick is 5 inches away. Remember that the binocular state actually does carry information about distance from the head.

A less simple and less plausible instance arose in the case of the confusion between Henry and George. In that case, we had to attribute a belief in the identity of the values of parameters of two nonsimultaneous states. As we noted, this belief really requires the supprt of other beliefs, especially when the tracking of the object is interrupted, even if only briefly. The tracking scenario, if developed, would have sharpened this latter point. Must we attribute such beliefs to our tracking frogs?

Let's return to Perry. Perry, [4], has been discussing a case in which he reaches out for a glass of water with the aim of quenching his thirst.

> Let us first note how unrealistic it would be to suppose that the content of our beliefs fix all of the circumstances relevant to the success of our action. Consider the force of gravity. If I am in space or on the moon or in some other situation where gravitational forces are much diminished, the movement we envisage me making will not lead to getting a drink; the water would fly out of the glass all over my face – or perhaps I would not even grab the glass, but instead propel myself backwards. If all possible failures are to be accounted for by false beliefs, then corresponding true beliefs must be present when we succeed. So, when I reach for the glass, I must believe that the forces of gravity are just what they need to be for things to work out right. But it hardly seems probable that everyone, even those with no knowledge of gravity, believes, when they reach for a cup of water, that the gravitational forces are what they are; such an attribution would drain the word 'belief' of much of its content.

Needless to say, the improbability becomes absurdity when applied, as it would have to be, to our frogs–both our real frogs and the imagined trackers. (After all, things don't just float around willy-nilly, unconstrained by gravity–neither frogs nor flies.) The more natural supposition is that evolution has resulted in an adaptation of frog psychology to various stable, essentially constant features of the environments in which a frog must operate. Some of these, like gravity, are both constant and non-species-specific, although the particulars to be accommodated and the modes of accommodation are specific. Others, such as regularities involving flies and their patterns of flight, are of interest only to some of Nature's creatures. Such regularities form precisely that background against which the purely psychological laws, involving beliefs, desires, and bodily *movements*, find their proper place. From the standpoint of design, a good designer would incorporate adaptability to these constant environmental features directly into the organism and worry later about what would happen if the organism found itself in a significantly different environment.

A.13 A Brief Digression on Defaults

Let us imagine that the bodily movement is much more expensive for the frog than it actually is; imagine that the movement is somewhat analogous to the sustained rush of a big cat after its prey. In such a case, we might want to characterize the story of "what the frog's eyes tell the frog's brain" in terms of a *default* or *ceteris paribus* rule: if an object causes such and such a pattern of ocular irradiation, assume it is a live fly and go after it. Now this is a defeasible presumption, of course, and there are many possible factors that would defeat it. Still, it is by and large a reliable guide to elief and action. In normal circumstances it leads to appropriate action. How should we think of this? Well, one alternative is as follows: there is a wide range of ways in which the environment might be abnormal and, moreover, be so in a way that would defeat the rule. The organism must believe, for each of those ways, a proposition to the effect that the environment is not abnormal in that way. That is, [the meaning of] its mental state must be extremely complex. This seems crazy. There was, after all no need to postulate a belief, on Kermit's part, in the regularity

connecting patterns of ocular irradiation with the presence of flies, nor in the regularitry connecting Ψ-type movements with the gobbling up of flies. If we need not attribute belief in these *ceteris paribus* regularities, need we attribute beliefs to the effect that things are, as a matter of fact, not abnormal in any of the relevant ways?

The aim of design in this regard can be put positively as follows: *jump to reasonable conclusions unless you shouldn't.* What a designer wants to do is to build the organism in such a way that, to put it crudely, the thought of its environment's being abnormal in any one of the relevant ways never crosses its mind *unless it should.* The aim is to design the creature in such a way that either being in the relevant cognitive state does not lead to the characteristic bodily movements in the presence of defeating abnormalities, or being in an abnormal situation blocks the production of the relevant control state. This is not a matter of miraculously transcending or bypassing the causal order. All of the relevant psychological and psychophysical laws operate against a set of background regularities and facts of various kinds. All that is required is that the organism somehow be sensitive to variations along the relevant dimensions in either of the two ways suggested above. So, for instance, the organism is to be built in such a way as to make it sensitive, say, to deviations (i.e., those that affect performance) from the envisaged norm in the local gravitational field. But its controlling mental state need not be such as to involve a component that represents that norm. In other words, that state does not have to involve a component substate whose meaning might be expressed as "The local gravitational field is within the normal bounds." (Or "The local gravitational field is as given by the following field equations.") Once again, there is a limit to what can be achieved in making an organism sensitive to significant abnormalities. But we shall get to the question of limits later.

A.14 What Must an Agent Believe (cont'd)

We have been discussing this topic at the level of the species, but now we should return to the imagined difference in tongue length between Kermit and Prince. Here it is ontogeny not phylogeny that is at issue; we are at the level of individual differences. One way to make up for the discrepancy between thought and action, desire and performance, is to suppose that our frogs have beliefs about the lengths of their respective tongues. In the case at hand, we might suppose that Prince has the false belief that his tongue is at least 5 inches long; if that belief were true, the bodily movement he executes would result in capture. Its falsity explains Prince's failure. Kermit, of course, has the true belief that *his* tongue is at least 5 inches long, or perhaps he believes that it is 6 inches long. The truth of *that* belief guarantees the success of his action in the circumstances. (This is the elegant, anthropomorphic way of expressing the earlier suggestion that Nature should so design things that flies a given distance away from Kermit and from Prince would appear to these unequally endowed frogs to be at different distances from them. In other words, Nature should try to make frogs immune from such a false belief as the one that makes Prince fall short.) Well, that is one way to go. But remember that there is another. We need only suppose that (at least within certain limits) individual psychologies are accommodated to individual physical differences. That is, the laws connecting perceptual states, desires, and

bodily movements in frogs like Kermit in regard to tongue length are a little different from those connecting the *same* perceptual state, desires, and bodily movements in frogs of lesser reach, such as Prince.

A.15 Intentional Realism Versus the Hypothesis of A Language of Thought

The general issue under discussion is whether the belief aspect of mental states **must** be such as to determine propositions whose truth would mediate fully between the bodily movement caused by being in that state and the successful attainment of the state's desire component. Our answer is: **NO**, they need not. But over and over we have suggested that there are limits. We hypothesize that only when we come up against such limits does the doctrine of a *propositional* language of thought underlying rational behavior receive any support. In particular, it is only when we come up against such limits that it *seems* necessary to postulate mental states (i.e., control states) that have as constituent parts tokens of sentencelike entities, that is, entities that are instances of structures from a (typically) infinite and inductively specifiable set. The alternative view we have been sketching is a version of what has been called *intentional realism.*

Let us remind the reader of what a *sophisticated* version of *intentional realism* provides:

1. It provides meaningful states whose instances bear information. Such instances represent various parts of the world as having certain properties and as standing in certain relations.

2. Moreover, it provides control states, i.e., meaningful states whose instances also cause bodily movements and whose meaning can be said to control those movements and rationalize the resulting actions.

3. Finally, these states are structured; indeed, these states have *semantically significant structure.* Various aspects or components or parts of these states are correlated, on the one hand, with various aspects or parts of the environments in which they are produced and, on the other hand, with various aspects or features of the bodily movements they cause. In even fairly simple cases, one can speak of certain aspects of such states as parameters of meaning and of the instances of such aspects within state instances as standing for the values of those parameters–that is, for individuals, properties, and relations.

   Thus, for example, we might imagine a simple creature able to perceive that a certain object is to the left of another (to the left, that is, from its own perspective.) Here the meaning of the relevant state might be characterized as follows:

   $$\alpha, l, o_1, o_2 \; [[\, \Xi \,]] \; o_1 \text{ is to the left of } o_2 \text{ at } l \text{ (from } \alpha\text{'s point of view).}$$

   Obviously, therefore, if an instance of Ξ can carry the information that Henry is to the left of Dick at l_1, then another instance can carry the information that Dick is

the left of Henry. Further, instances can carry the information about any two objects that α can perceptually discriminate that one is to the left or, *for that matter, to the right, of the other*. Hence, these states exhibit some of the productivity or systematic nature characteristic of–but not, we contend, distinctive of or unique to–language.

Now what more is needed or wanted? One possible thought is that there are mental states that have aspects that denote other such states. Or, which *might* be the same point, that some states have structural components that are themselves meaningful states. In other words, some states actually involve, as parts, tokens (instances) of other states. This is one interpretation of the claim that the so-called propositional attitudes (e.g., beliefs and desires) are to be understood on the model of relations to sentences. Now of course if Kermit had beliefs about his own or Prince's (or Henry's) beliefs, we might have to face the problem of aspects of meaningful states denoting meaningful states. And people certainly do have such beliefs; but of course people also use genuine natural languages–English, Japanese and the like. Is the use of a natural language a necessary condition of the possibility of having beliefs about one's own or another's mental states? Probably not; but the issue is too difficult to allow of brief treatments.

A.16 The Question of Limits

Is there *another* way of getting at the intuitions behind the doctrine of a language of thought? There is and, moreover, we think we can see what it is by returning to the question of limits. We've tried to make plausible the claim that accommodation to or control by certain types of information that are relevant to action can be achieved without postulating *beliefs* whose content comprises those types of information. The types of information have involved the following:

- General, stable characteristics of the environments within which members of the relevant species operate, general stable characteristics of those members, and general, stable regularities connecting chracteristics of the species and their environments.
- Specific bodily and physiological characteristics of individual organisms.
- Perceptually given facts about certain (perceptible) aspects of the current environment.

Now what about such a fact as that Henry was near Palo Alto, California, on May 31, 1986, at such and such a time? This is not a fact (not a piece of information) that belongs to any of the foregoing three types. It *is* a piece of information that some state of Kermit can bear; indeed, an instance of Φ could carry it, as long as that instance occurred in Kermit while he was near Palo Alto and the instance was caused by Henry's flitting about. But how can Kermit's ability to carry this information (that is, how can the fact that Kermit can be in a state which carries this information) be of any use to Kermit in controlling fly-directed behavior? Now it's time to return to the fiction of the tracking frogs. And, of

course, it is time to drop any fiction with respect to the significance of Palo Alto in the mental lives of frogs. Let us substitute for Palo Alto a particular hydrangea h_1–a red one, of course. We're going to be imagining a world in which flies are drawn to red hydrangeas as if they were the sweetest honey. Now for more fictional ethology, still on the level of general facts about normal environments. Red hydrangeas are quite scarce and widely dispersed in Kermit's environment and so are flies. Moreover, the flying range of flies is fairly limited. We'll have to fictionalize a good deal here anyway, so we shall assume, for instance, that frogs can discriminate among various flowers and colors. But we shall *not* assume that frogs are unconditionally disposed, when hungry, to "gravitate toward" red hydrangeas.

Let us suppose that Kermit sighted Henry near h_1 at some time t_1 earlier today and that he then continued without interruption to track Henry visually. Despite his vigilance, however, he lost sight of the little bugger at t_2. Now what? Well, Kermit at t_1 *was* in a state that carried the information that Henry was near h_1 at t_1. That's a useful bit of information, especially given the average fly's tendency to head for red hydrangeas, together with general facts about the scarcity and wide dispersion of these in the flies' environment. But how can Kermit make use of it?

Well, first, what action would it make appropriate? That depends on what capabilities we are ascribing to our frogs. According to one, relatively modest account, it is stopping and waiting at h_1 (but not at h_2, another red hydrangea that is quite remote, as the fly flies, from h_1.) So the type of ***bodily movement*** in question, call it Σ, involves (on this more this more modest account) something rather like the UNTIL construct of programming languages: wander about until entering a perceptual state of type Θ, where Θ is the characteristic perceptual state caused by seeing or smelling a red hydrangea. But note: not until perceiving h_1; *individuals do not enter into the meanings* either of mental states or of bodily movements. Still, we could have told this story slightly differently, with the problem involving ***recognizing*** h_1 when Kermit encounters it again. That capacity, indeed, might seem to be implicit in the specification of the desired ***action***: wander around till you're back at h_1 and stop. We've cheated a bit here and assumed that red hydrangeas are relatively few and far between within the frog's immediate environment. No matter, we'll be opting for a more ambitious account anyway. Also, we take it for granted that a sighting of a fly in flight acts as a ***hard interrupt***. The frog is built to drop everything and visually track its prey (till capture, if possible) as long as it's hungry enough. Just as this general fact about the species need not be the content of any mental state on any frog's part, the specifications of all those bodily movements that play a part in the search for food need not be complicated by an aspect whose meaning is as follows: if in state Δ, drop everything and start tracking–where the meaning of Δ is a generalization of the meaning of Φ.

Another, more ambitious and complicated account, the desired action is to hike (hop?) back to that red hydrangea where you spotted Henry, that is, to get back to h_1. On either the modest or the more ambitious account, the problem for Kermit can be put as follows: the proposition that Henry was at h_1 at t_1, the proposition that would be good for Kermit, at t_2, to believe in a bodily-movement-controlling kind of way, is not a part of the circumstances in place at t_2. It is not one of those facts about c_2 that mediate

between Kermit's bodily movement and his getting back to where he sighted Henry and to which place Henry is liable to return. (Nor, of course, is it a fact about Kermit's particular physiology or physiognomy.) In any event, Kermit must store the information his state carried at t_1 and must do so in a way that allows for the control of behavior (at t_2).

Let's go along with the ambitious account. This account brings in quite explicitly some notion like that of a *mental map*; because of this, it is both a more compelling and complicated story, as well as more revealing of the roots of the doctrine of a language of thought. Consequently, a word on mental maps is in order here to help drive home the point we have been gradually getting to, albeit circuituously. Here it will be best to consider real atlas-type maps as well, so let us now cease talking of frogs and talk, instead, of people.

A.17 Maps

There are many kinds of mental maps. The crucial differences among them, for our purposes, lie along the single dimension of context-relativity and agent-centeredness. We shall simplify and speak of two varieties. One is highly agent-centered; the information such maps carry is very closely connected with the contexts and circumstances in which they are produced. The mental maps we have in mind here are the so-called visual and auditory maps in the respective cortices and the body maps in the somatosensory cortex. These maps are akin to states like Φ and Ψ and have little in common with atlas-type maps. Mental maps of the other type, like atlas-type maps, are less agent-centered; the information they bear is much less determined by the circumstances in which the states occur. Within broad and somewhat indeterminate limits, real maps can be used by anybody, anywhere, in any circumstances, as long as the agent knows how to read them and how to orient himself with respect to it. Think of a real map of California–even one you've committed to memory. Think of how you use it. You have to place and orient yourself on it. The map has no special aspect for you and your current location and direction. This *you* must supply, along with the connection between your present whereabouts and the map. If you can do these things, you can use the map. Of course, you can come to know your location and orientation with respect to a map by hypothesizing values for these parameters and testing the hypothesis. Notice that you can use it even if you are located outside the area covered by the map. In such cases, you have to be able to orient yourself with respect to the map, but not to place yourself on it.

The information that atlas-type maps carry is specific; it is about particulars. It is not about (1) general characteristics of the environments within which members of the relevant species typically operate. It has nothing to do with (2) the specific bodily and physiological characteristics of individual members of the species. Finally, and crucially, the information a map bears, including the information about what may in fact be an organism's current environment, is not (3) perceptually available to that organism in a behavior-controlling way.

[Atlas-type] maps carry content in a highly context-independent manner; the information they carry is not in a form that is **directly useful** for control of behavior. **That is, by themselves, inside the head or out, they are useless for the control of behavior.**

But, together (and only so) with the ability to orient oneself with respect to them, they allow us to *use* information that, without them, could be borne by our states, but only uselessly. That is, they allow us to employ, in the control of behavior, information (propositional content) that goes beyond what is otherwise available to us in the circumstances in which we move and that, moreover, is required if those movements are actually to promote our goals.

Here you are (unbeknownst to you) in Palo Alto. You want to walk to Mountain View. Simplifying the procedure considerably, let us say that the appropriate bodily movement for you is to walk in a southward direction. Nothing in your current perceptually accessible circumstance tells you that you are in Palo Alto, but you do see a beautiful building of a distinctive shape in front of you. You get out what you know to be a detailed landmark map containing your present whereabouts (never mind how you know this). The only place name on the map is Mountain View; moreover, the map has no explicit directional markings. You therefore orient yourself on it by means of the one symbol on the map that is conceivably a symbol for the beautiful building before you. Perhaps you turn yourself, or the map, around so that, as you look at the map, you're also looking in the direction of Mountain View, with that beautiful building at your back. And off you go, southward.

What beliefs do you have to have to make sense of this last bit of behavior? That you are now directly to the north of Mountain View, facing south? But that's not what the map told you. What it *did tell* you–once you found your place and orientation with respect to it–was: *here* you are and *there* is where you want to get to, off in *that* direction–straight ahead. *This is* what guides your behavior and *it* is highly context-sensitive. On the other hand, the proposition that must be true if your walking straight ahead, in the circumstances envisaged, constitutes walking toward Mountain View (or is to advance your goal of walking to Mountain View) is that, from where you are now, i.e., Palo Alto, facing in the direction you're facing, i.e., southward, Mountain View is straight ahead. (Put more simply, the relevant proposition is that you are in Palo Alto, facing south, and Mountain View is south of Palo Alto.) The information conveyed by the map thus mediates between controlling aspects of the circumstances in which bodily movement occurs and the propositions that must be true for that bodily movement, executed in those circumstances, to constitute advancement of the agent's goal. It is in this way that maps can act as stable, context-independent links between control states, but they are not themselves, and by their very nature cannot be, the meanings of control states.

The information conveyed by maps is obviously very handy. It allows for the control of behavior by states carrying information about specific facts that are quite remote from the circumstances in which that behavior is to be produced. It thus allows for a degree of foresight and *planning*. Maps can be used to get where you want to go, even in those cases in which, perhaps because of sheer distance, your current circumstances yield no perceptible clues to the relative location of your intended destination.

Let us return, briefly, to our old friend Kermit. Imagine that he has a nonagent-centered mental map of his little environment with a landmark for h_1 and other relevant salient sites. If he can use this map and can find his present position and orientation with respect to

it (by recognizing that he is facing some nearby marked site), he can have some hope of directing his movements over a fairly long stretch of time, during which there will almost certainly be periods when his local circumstances will offer him no clue as to the relative location of h_1. This would give him and his ilk an enormous advantage in terms of natural selection.

As we've noted, however, such information, if it is to be truly useful, must be stored in a way that makes it possible for the creature to orient itself with respect to it. That is, the creature must be able to tell where and when (there are also such things as temporal maps, which might be called *schedules*) it is situated in terms of the coordinate system on which the map is based. Does this mean that, when performing actions, the creature must be able to *read* these mental maps, as opposed to having its behavior controlled and, in part, caused by them? No, but it does mean that *we theorists* must be able to provide some account about storage and retrieval of such relatively context-independent bearers of information, as well as about an organism's ability to orient itself with respect to them.

We are not going to try to provide such an account here. Let us note that it begins with the following kind of situation: an organism has two [or, at any rate, a small number of] relatively stable objects in its perceptual field at a certain time. Its visual state carries the information that o_1 is to the left of o_2. Moreover, its visual and body maps are coordinated in such a way that we may also characterize the organism's total state as follows: if it wanted to grab at o_1 it should reach out its left arm (or paw). Now the organism turns around, say, to its right, for a while retaining some peripheral view of both objects, then losing sight of them, first of o_1, then of o_2. (Of course, we may have to take into account the fact that the organism cranes its neck to the left, so as to keep both objects in sight as long as possible.) It ends up with its back to the two objects, receiving no visual stimulus from either. As it moves its body and head (and eyes), its visual and bodily maps are automatically *transformed and recoordinated* so that its state at the later time, while carrying information about the objects then in its visual field, will also convey the information that the two objects are behind it and that o_2 is now on its left.

We may know how to realize such a capacity in robots, without having a clear idea as to how Nature does it. In any event, one should note that some information about the relative locations of objects can be borne by context-sensitive states of organisms–and in a form that is suitable for control–even when such objects are no longer in sight. The point here is to contend that, in explaining this kind of phenomenon within the framework of intentional realism, as we also *claim* can be done, we shall also be accommodating the central intuitions that support the doctrine of a propositional language of thought.

References

[1] K.J. Barwise, *Information and Circumstance*, **Notre Dame Journal of Formal Logic**, vol. 27, no. 3 (July), pp. 324-338.

[2] J.A. Fodor, **Psychosemantics: The Problem of Meaning in the Philosophy of Mind**, unpublished ms.

[3] J.R. Perry, *Self-Knowledge and Self-Representation*, in **Proceedings of IJCAI-85**, Morgan Kaufman Publishers, Los Altos, CA., pp. 1238-1242.

[4] J.R. Perry, *Circumstancial Attitudes and Benevolent Cognition*, in **Language, Mind and Logic**, J. Butterfield, ed., Cambridge University Press, Cambridge, England.

[5] J.R. Perry, *Thought without Representation*, unpublished ms. of a lecture presented at a Joint Symposium of the Mind Association and the Aristotelian Society, London, July, 1986.

[6] B.C. Smith, *Varieties of Self-Reference*, in **Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning about Knowledge**, J. Y. Halpern, ed., Morgan Kaufman Publishers, Los Altos, CA., pp. 19-43.

APPENDIX B

By Robert C. Moore

B Events, Situations, and Adverbs

B.1 Introduction

This report concerns a dispute about the relationship of sentences to the events they describe and how that relationship is manifested in sentences with adverbial modifiers. The two sides to the argument might be called the "Davidsonian position" and the "situation semantics position", the former being chiefly represented by Donald Davidson's well-known paper "The Logical Form of Action Sentences" [1] and the latter by John Perry's critique of Davidson's view, "Situations in Action" [2].[3]

The issue turns on Davidson's analysis of how a sentence like (1) is related to a similar sentence with an adverbial modifier, such as (2):

(1) Jones buttered the toast.

(2) Jones buttered the toast in the bathroom.

Stated very informally, Davidson's position is that sentence (1) claims that an event of a certain type took place, to wit, a buttering of toast by Jones, and that (2) makes a similar claim but adds that the event took place in the bathroom. Put this way, an advocate of situation semantics could find little to complain about; Perry and Barwise themselves say rather similar things. The dispute is over the way in which (1) and (2) claim that certain events took place. Davidson suggests that the event in question is, in effect, a hidden argument to the verb "butter". As he would put it, the logical form of (1) is not

(3) *Buttered(Jones, the toast)*

but rather

(4) $\exists x$*(Buttered(Jones, the toast, x))*,

where the variable x in (4) ranges over events.[4] Adding the adverbial modifier is then quite straightforward; it is simply an additional predication of the event:

(5) $\exists x$*(Buttered(Jones, the toast, x)* $\wedge$ *In(the bathroom, x))*.

Perry objects strenuously to making the event described by the sentence an explicit argument to the relation expressed by the verb. He says:

[3]This dispute is really just a special case of a much deeper disagreement about semantics that is treated in depth by Barwise and Perry in *Situations and Attitudes* [3].

[4]This analysis obviously makes no attempt to analyze the tense of the verb or the structure of the noun phrase "the toast".

> If we ask what about the statement tells us that there was an event of that type, the only reasonable answer is that the whole statement does. It is not that part of the statement refers to an event, and the other part tells us what it was like. Part of the statement refers to Jones and the other part tells us what he did. Both parts working together tell us that an event of a certain sort occurred. The simple parts of the sentence refer to basic uniformities across events: Jones, buttering, and the toast. The way the simple parts are put together in the sentence describes the event. [2, p. 2]

Davidson considers but rejects an analysis derived from Reichenbach [4, pp. 266–274] that is in the spirit of Perry's objection. In this analysis, (1) and (2) would be rendered by (6) and (7), respectively:

(6) $\exists x$*(x consists in the fact that Jones buttered the toast)*

(7) $\exists x$*(x consists in the fact that Jones buttered the toast and x took place in the bathroom).*

This seems to meet Perry's objection in that it is the whole statement "Jones buttered the toast" that gives rise to the reference to the event, rather than a hidden argument to the verb. Davidson rejects the analysis, however, on the grounds that its logical properties are problematical. Davidson notes that from the identity of the Morning Star and Evening Star we would want to be able to infer that if I flew my spaceship to the Morning Star, then I flew my spaceship to the Evening Star. On the analysis under consideration, this requires being able to infer (9) from (8).

(8): $\exists x$*(x consists in the fact that I flew my spaceship to the Morning Star)*

(9) $\exists x$*(x consists in the fact that I flew my spaceship to the Evening Star).*

Davidson argues that the only reasonable logical principles that would permit this inference to go through entail the identity of *all* actually occuring events, which would be absurd. Barwise and Perry's [3, pp. 24–26] rejoinder is that Davidson makes the unwarranted assumption that logically equivalent sentences would have to be taken to describe the same event, an idea they reject. Perry [2] then develops, within the framework of situation semantics, an analysis of event sentences and adverbial modification that is faithful to the idea that, in general, it is an entire sentence that describes an event.[5]

To summarize the state of the argument: Davidson and Perry agree that sentences describe events, but Davidson thinks that it is virtually incoherent to view the event as being described, as it were, "holistically" by the entire sentence, whereas Perry views it as "the only reasonable answer." Barwise and Perry pinpoint where they think Davidson's argument goes wrong, and Perry provides an analysis of adverbial modification consistent with the holistic view.

[5] We omit the details of Perry's own analysis of adverbial modification, as it would require a far more extensive presentation of the ideas of situation semantics than is really appropriate for this paper, and it is not really needed for the points we wish to make.

B.2 Some Facts About Adverbs and Event Sentences

Perry's and Davidson's analyses have in common that neither is based on a very extensive survey of the *linguistic* data to be accounted for by a theory of adverbial modification. The strongest considerations motivating their analyses are more general logical and metaphysical concerns. A more careful examination of the relevant linguistic phenomena, however, shows that neither Davidson nor Perry have the story quite right, and that a more complete account of adverbial modification has to make room for at least two views of the relation between sentences and events, one close to Davidson's and the other close to Perry's.

The key set of data that I will try to account for is a significant class of adverbs that can be used to modify event sentences in two quite distinct ways:

(10) (a) John spoke to Bill rudely.
(b) Rudely, John spoke to Bill.

(11) (a) John stood on his head foolishly.
(b) Foolishly, John stood on his head.

(12) (a) John sang strangely.
(b) Strangely, John sang.

The difference between the first and second member of each pair should be clear. For instance, (10a) suggests that it was the way that John spoke to Bill was rude, while (10b) suggests that the very fact that John spoke to Bill was rude. Thus (10a) leaves open the possibility that John could have spoken to Bill without being rude, but (10b) does not. Similar remarks apply to the other pairs; with this class of adverbs, in general, "X did Y Adj+ly" means that the way X did Y was Adj, and "Adj+ly, X did Y" means that the fact that X did Y was Adj. We will therefore say that the (a) sentences involve a "manner" use of the adverb and that the (b) sentences involve a "fact" use.

A significant difference between the fact and manner uses of these adverbs is that the manner sentences are extensional with respect to the noun phrases in the sentence, whereas the fact sentences are not. That is, we may freely substitute coreferential singular terms in the manner sentences, but not the fact sentences. Suppose it is considered rude to speak to the Queen (unless, say, she speaks to you first) and suppose John is seated next to the Queen. Then it could well be that (13) is true, while (14) is false, although they differ only in having substituted one singular term for a coreferring one:

(13) Rudely, John spoke to the Queen.

(14) Rudely, John spoke to the woman next to him.

Sentence (14) can differ in truth-value from (13) because—on at least one interpretation—(14) seems to entail that it was rude for John to speak to the woman next to him, *whoever* she was, i.e., even if she were not the Queen. The issue is somewhat complicated by the fact that these sentences seem to exhibit the sort of *de dicto/de re* ambiguity common to most nonextensional constructs. That is, (13) and (14) seem to be open to an an additional interpretation, whereby we might be saying that there is a certain woman, whom we may

identify either as the Queen or the woman next to John, and that it was rude for John to speak to that particular woman.

On the other hand, it seems that (15) and (16) must have the same truth-value on *any* interpretation, so long as the Queen and the woman next to John are the same person. Moreover, no *de dicto/de re* distinction seems to obtain:

(15) John spoke to the Queen rudely.

(16) John spoke to the woman next to him rudely.

Note, however, that (15) and (16) are not completely extensional in the sense that first-order logic is extensional. That notion of extensionality requires not only intersubstitutivity of coreferring singular terms, but also intersubstitutivity of sentences with the same truth-value. But even if (17) and (18) have the same truth-value, it does not follow that (19) and (20) do:

(17) John spoke to the Queen.

(18) John spoke to the Prince.

(19) John spoke to the Queen rudely.

(20) John spoke to the Prince rudely.

This sort of behavior is quite general with these adverbs. Examples similar to (13)-(20) can be constructed for "foolishly," "strangely," and all the other adverbs in this class.

A last set of observations, pointed up by our use of "Adj+ly" and "Adj" in the schematic statment of the generalizations given above, concerns the fact that the adverbs in the class we are considering typically have related adjectival forms:

(21) It was rude for John to speak to Bill.

(22) It was foolish for John to stand on his head.

(23) It was strange for John to sing a song.

We note that these adjectives can all take clausal complements (in the case of (21)-(23), infinitives), and that they are all *factive*. That is, for "It was Adj for X to do Y" to be true, it must be the case that "X did Y" is true. We also note that such adjectives can be applied to noun phrases that refer explicitly to what seem to be events:

(24) John's action was rude.

(25) John's mistake was foolish.

(26) John's performance was strange.

This suggests that the idea, common to Davidson and Perry, that adverbs have some important connection to events is on the right track. Exactly what that implies, we will see in the next section.

B.3 States of Affairs and Events

What, then, are we to make of this data? The short answer is that Davidson's analysis seems to give the best picture of the manner use of adverbs, while something close to Perry's seems to give the best picture of the fact use.

The analysis depends crucially on the notion of what we will call a "state of affairs". This term has been used by various writers for various notions, so we will go into some detail about the notion used here. It will be useful to contrast states of affairs with propositions, because it might seem at first glance that one could dispense with states of affairs in favor of propositions in doing the semantics of natural language. We will see that this is not the case.

A word of caution is in order before proceeding further. The goal of this exercise is semantical analysis of natural language, not the discovery of Deep Metaphysical Truths. If we postulate states of affairs as entities in the world, it is not because we believe that such things are really out there, but because postulating them gives the most natural analysis of the meanings of the class of sentences we are trying to analyze. The issue thus is what metaphysics is embedded in the language, not whether that metaphysics is actually true.

We will begin drawing the contrast between propositions and states of affairs by noting that English provides ways of referring to them explicitly, but with different types of construction. Propositions are typically denoted in English by "that" clauses, while states of affairs are typically denoted by gerunds:

(27) Mary believes *that John is foolish.*

(28) *John's being foolish* cost him his job.

Sentence (27) says that Mary believes a certain proposition, the proposition that John is foolish, while (28) says that a certain state of affairs, John's being foolish, cost John his job. The correlations between propositions and "that" clauses and between states of affairs and gerunds are not perfect. Sometimes "that" clauses also seem to denote states of affairs, and, as we shall see later, gerunds can denote events, which are also to be distinguished from states of affairs. These correlations, then, do not give foolproof criteria for recognizing reference to propositions or states of affairs, but they are useful in generating examples for consideration.

Roughly speaking, we take a state of affairs to be a particular, existing condition in the world. We would want to say, for example, that if the cat is on the mat, then the cat exists, the mat exists, and a state of affairs of the cat's being on the mat exists. For us, then, states of affairs are concrete particulars. This is important, because often the term "state of affairs" is used to describe abstract entities which may obtain or not obtain, as in the locution "The state of affairs of the cat's being on the mat does not obtain." For us, these would be *types* of states of affairs rather than states of affairs themselves.

This leads to one of the most important distinctions between propositions and states of affairs: While every meaningful sentence expresses a proposition (at least in a particular context), only true sentences describe states of affairs. Consider how the sentence "Reagan is a Democrat" relates to the following two examples:

(29) John believes that Reagan is a Democrat.

(30) Reagan's being a Democrat accounts for his foreign policy.

Sentence (29) is relatively unproblematic. It merely asserts that John believes a certain proposition, the propostition that Reagan is a Democrat. The fact that this proposition is false does not matter one way or the other in analyzing the meaning of the sentence. It purports to say that a certain state of affairs, that of Reagan's being a Democrat, accounts for Reagan's foreign policy, but there is no such state of affairs. The sentence fails in exactly the same way as does Russell's famous example:

(31) The present king of France is bald.

One may wish to say that (30) and (31) lack truth-values or that they are false, according to one's theory of sentences containing nonreferring terms.

References

[1] D. Davidson, *Essays on Actions and Events*, Essay 6, pp. 105–148 (Clarendon Press, Oxford, England, 1980).

[2] J. Perry, "Situations in Action," unpublished ms. of a lecture presented at the annual meeting of the Pacific Division of the American Philosophical Association, March 1983.

[3] J. Barwise and J. Perry, *Situations and Attitudes* (The MIT Press, Cambridge, Massachusetts, 1983).

[4] H. Reichenbach, *Elements of Symbolic Logic* (Macmillian Co., New York, New York, 1947).

END

7-87

DTIC